万卷方法 | 统计分析方法丛书

PEARSON

如何解读统计图表：研究报告阅读指南

Interpreting Data: A Guide to Understanding Research

彼得·M.纳迪（Peter M. Nardi）著

汪顺玉 席仲恩 译





重庆大学出版社

http://www.cqup.com.cn

RUHE JIEDU TONGJI TUBIAO：YANJIU BAOGAO YUEDU ZHINAN

■ 本书专门讲解如何阅读和理解报纸、大众杂志及学术期刊中常用的一些统计图表。实际上，对于那些想要学习统计知识和图表分析，而又觉得传统教材枯燥而艰深的人而言，阅读本书将有助于他们更容易、更迅速地掌握基本的统计分析技术，感受并发现统计分析的趣味和价值所在。

■ 作为大学老师，本书作者纳迪教授先后讲授过13门课程，其中有“定量研究方法”“计算、计算机和社会”“统计学入门”“沟通社会学”等。作为大量学术著作和科学普及著作的作者、沟通学教师和通俗讲座人，他有着高超的学术沟通艺术，擅长把抽象的学术概念和内容用通俗的日常生活语言表达，这使得本书生动而易读。

发表及参阅相关讨论，请登录：

万卷方法与学术规范博客圈（http//q.blog.sina.com.cn/fafang）


ISBN 978-7-5624-4906-5

定价：18.00元

万卷方法 | 统计分析方法丛书

如何解读统计图表：
研究报告阅读指南

Interpreting Data:
A Guide to Understanding Research

彼得·M.纳迪（Peter M. Nardi）著
汪顺玉 席仲恩 译

重庆大学出版社

图书在版编目(CIP)数据

如何解读统计图表:研究报告阅读指南/(美)纳迪著;汪顺玉,席仲恩译.—重庆:重庆大学出版社,2009.8
(万卷方法.统计分析方法丛书)
书名原文:Interpreting Data: A Guide to Understanding Research
ISBN 978-7-5624-4906-5

Ⅰ.如… Ⅱ.①纳…②汪…③席… Ⅲ.统计图—识图法 Ⅳ.C815

中国版本图书馆 CIP 数据核字(2009)第104606号

如何解读统计图表:研究报告阅读指南

彼得·M.纳迪 编
汪顺玉 席仲恩 译
责任编辑:林佳木 谢 芳 版式设计:林佳木
责任校对:文 鹏 责任印制:赵 晟

重庆大学出版社出版发行
出版人:张鸽盛
社址:重庆市沙坪坝正街174号重庆大学(A区)内
邮编:400030
电话:(023) 65102378 65106042(24小时录音电话)
传真:(023) 65103686 65105565
网址:http://www.cqup.com.cn
邮箱:fxk@cqup.com.cn (营销中心)
全国新华书店经销
重庆升光电力印务有限公司印刷

*

开本:940×1360 1/32 印张:5 字数:129千
2009年8月第1版 2009年8月第1次印刷
印数:1—3 000
ISBN 978-7-5624-4906-5 定价:18.00元

译者前言

当今世界是一个充满着数据的世界，不仅各种学术期刊、年鉴和报告中充满了数据，各种媒体报道和资讯中也充满了数据。如何解读这些数据，从这些数据中提取有用的信息，从而充分地利用这些信息进行知情决策，这不仅是每一个科学（包括自然科学和社会科学）工作者必须面对的问题，也是每一个生活在今天这个信息爆炸时代的普通人所不可避免和应该直面的问题。可见，解读数据不仅是每一个科学人必备的基本职业技能之一，也是每一个社会人为了更好地生存所应具备的基本生活技能之一。

统计学的原理可以帮助我们理性地采集数据，有效地处理数据，正确地解读数据。可是，初次学习统计学原理的社会科学工作者和学生大都有这么一种感觉：统计学原理过于空洞，学起来如坠云雾之中。这种感觉不仅我们中国的学者和学生有，美国及其他国家和地区的学者和学生也都有。尽管这种感觉是由该学科的特有性质造成的，有这种感觉也很正常，但是，这种感觉却是我们在学习中必须努力克服的，应该被踏实感和实在感取而代之。因为，踏实感是我们对统计原理和技术有把握、有信心的标志，是我们学懂了、放心了、敢用了的自我心理反应；实在感是我们知道怎样去理解、怎样去学习、怎样去操作的心理反应。

为了使广大初学人员对统计学这套方法和技术有一种踏实感，人们才编撰出了各种各样的统计学教材，开设了各形各色的统计学课程；为了使初学统计学原理的人们在学习过程中有一种实在感，彼得 · M. 纳迪博士才编写了这本《如何解读统计图表：研究报告阅读指南》。显然，这是一本基础统计学课程的教学辅助性读

物,是一般基础统计学课本的伴读。

彼得·M. 纳迪博士是皮兹尔学院(Pitzer College)的社会学教授,他曾任"太平洋社会学协会"(Pacific Sociological Association)2005—2006 年度的主席,担任过《社会学透视》(*Sociological Perspectives*)、《性学》(*Sexualities*)等学刊的编辑,担任过包括《美国社会学评论》(*American Sociological Review*)在内的 19 个学刊和包括 Sage 出版公司在内的 10 家出版社的审稿人,发表了大量社会学学术论文,编辑和编写了多部学术著作,其中两部(包括本书)是关于定量研究方法的。纳迪教授不仅致力于学术研究,而且非常关心科学普及工作。他曾为多部百科全书撰写词条,为多部著作写过专业书评。更难能可贵的是,他还做过将近八十场的专题讲座,有些是针对大学生讲的,有些是为普通听众讲的。

作为大学教师,纳迪教授先后讲授过 13 门课程,其中一门是"定量研究方法"(Quantitative Research Methods),一门是"计算、计算机和社会"(Computing, Computers, and Society),一门是"统计学入门"(Introductory Statistics),还有一门是"沟通社会学"(Sociology of Communications)。作为一名经验丰富的研究者和审稿人,纳迪教授非常了解社会学的学生应该掌握哪些定量研究方法和统计技术,特别是他们还欠缺哪些必备的方法和技术;作为一位资深的定量研究方法和基础统计学老师,他深谙学生初学统计学时的困难,并尝试过各种帮助学生排解困难的方法,而且获得了不少的一手经验;作为大量学术著作和科学普及著作的作者、沟通学老师和通俗讲座人,他有着高超的学术沟通艺术,擅长把抽象的学术概念和内容用通俗的日常生活语言表达出来,不仅自己表(述),还便于读者(抵)达。

是的,用通俗易懂的日常生活语言介绍和讲解抽象而复杂的统计学原理和技术,让读者一看就明白,是《如何解读统计图表:研究报告阅读指南》这本书最显著的特点之一。这本书的显著特点之二是拟真性和操作性俱强。翻看本书的任何一个章节,都可以一下子感受到,书中对示例数据的解读几乎和真实的研究过程一样。特点之三是示例内容的时代性、普通性以及数据的真实性。全书所用的示例都取自最近几年的真实调查结果,有些取自通俗

报刊,有些取自学术期刊,内容都是大家普遍关注的问题。例如,每天花在看电视上的时间、性教育问题等。特点之四是布局结构化。首先,每章都用同样的安排。每章都以对本章基本概念的定义开始,然后是有关本章内容的一些假定,接着便是解读示例,最后以本章小结结束。其次,示例的解读也是结构化的。每个示例统一按照"找出变量"、"解读图表"和"得出结论,解释结果"的版块顺序展开。再次,练习中的问题也基本是结构化的。每个练习题的第一个问题都是变量辨认和变量测量水平确定,最后一个问题是结果解释,其中还包括为什么该方法适合该问题以及用日常生活语言做研究结论等。这种安排看似单调,实则属于作者的精心设计。这种设计,对于初学统计学的学生尤为重要。因为这样可以使学生减少很多没有必要的陌生感和云雾感。虽然这本书还有很多其他方面的特点,有待各位读者见仁见智,但是,至少还有一点是显而易见的,那就是,本书中的数据和图表大都来自"研究检索"这个强大数据库。为了方便读者调用,书中提供了发表该数据/图表的文章的"研究检索索引号"。

下面,我们先简要介绍一下各章的内容,然后说明一下我们翻译中遇到的问题,最后说明一下本书的翻译过程和各位译者的具体分工。

《如何解读统计图表:研究报告阅读指南》一书的正文部分由内容"导论"和五个独立章构成。"导论"介绍了全书的安排,定义了以后各章的通用术语,特别是介绍了统计决策的流程图。五个独立章分别是关于基础统计学的一组统计技术的例解。

第 1 章:描述数据。主要解读几个常用的描述统计量和统计图,重点是各种常用统计图表的解读,包括饼图、条形图、折线图、频数(率)表。统计图表的功能是辅助性的。通过对调查结果数据的直观展示,能方便我们找出更为合适的统计技术。

第 2 章:理解表格。重点解读各种交叉表,包括简单的二元统计表和较为复杂的多元统计表。交叉表的主要功能是发现问题,而不是解决问题。"一旦你得到了一组结果,你的任务才刚刚开始"。

第 3 章:解释相互关系。这一章的重点是卡方检验和皮尔逊

积矩相关。关于积矩相关,应该特别注意的是:相关关系不等于因果关系。关于相关关系的量度参数皮尔逊 r 的解读,应该注意数值的大小和正负号。数值大小是变量之间关系强弱程度的量度,正负号是变量之间关系方向的标志。卡方检验只能检验变量之间有无关系,或者说变量之间的关系是否显著,但不能度量关系的强弱程度。皮尔逊积矩相关要求两个变量都在定距或定比水平上取值。对于定类变量和定序变量,卡方检验则是一种理想的方法。

第 4 章:解释均值差异。本章主要解读 t 检验和单向方差分析的统计结果。要比较两组观察值在均值上的差异是否显著,t 检验是理想的选择。要比较三组以上(包括三组)观察值在均值上的差异是否显著,方差分析是最佳的选择。t 检验和方差分析都要求因变量在定距或定比水平上取值。

第 5 章:解读回归数据。回归数据解读的要点是决定系数 R^2 和回归系数 B 与 β。决定系数告诉我们,因变量上的变异方差,有多大的比例能够被自变量上的变异方差所解释。回归系数表示自变量对于因变量的贡献大小和方向。比较不同自变量的贡献大小时,应该注意各自变量的单位。如果各自变量的单位相同,用回归系数 B 即可,如果各自变量的单位不同,则要看标准回归系数 β。

尽管《如何解读统计图表:研究报告阅读指南》一书是一本辅助性著作,我们在翻译中还是遇到了不少困难。首先,题目的翻译就很难处理。原书的题目是 *Interpreting Data: A Guide To Understanding Research*,既有主标题,也有副标题。如果直译,题目似乎应该是:《解读数据——了解研究之向导》。这种表述在汉语中显然是不通的。经再三商议,决定采用现在的译名。其次,术语翻译是让我们最为头疼的。如果我国和美国在应用统计学和研究方法方面的发达程度相当,就不会存在术语翻译的困难,至少会少很多。但事实上,我国和西方发达国家特别是美国在这两个学科上的差距,远远大于在数学、物理、化学、生物学等学科分支上的差距。单是 variable 这个词的翻译就很棘手。因为在汉语中很难找到一个合适的词,以涵盖这个英文词在该书中的意思。按理,汉语中的"(可)变(因)素"较为合适,但是国内却习惯于用"变量"。为了尽量不对读者造成没有必要的术语陌生感,我们虽不情愿,但最

后还是用了“变量”这一译法，尽管这个所谓的“量”在很多情况下并不能表示大、小、多、少等关系，或前、后、迟、早等顺序。

应用统计学难免要涉及一些研究方法和社会测量学方面的知识，而这两个分支又不是那么成熟，并且原文中的部分用语本身就比较随便，因此，这给翻译工作又增添了不少的困难。例如 measurement 一词，在原书的不同语境中就有不同的意思。最常见的两个意思是“测量过程”和“测量结果”。在科学技术和工程、质检等领域的计量学文献中，measurement 一词只允许有“测量过程”这一个意思。levels of measurement 是应该译成“测量水平”呢，还是译成“测量结果的水平”？尽管我们知道，严格的英语表述应该是 levels of quantity，对应的汉语应该是“量的水平”，即变量在哪个量的水平上取值。考虑到本书的伴读性，我们颇不情愿地做了折中处理，在一些情况下采用了“测量水平”，在部分可能给读者造成混乱的语境中采用了“量的水平”。在翻译“量”的各种具体水平时，我们决定绝不随大流，不折不扣，因为我们害怕随大流可能会导致很多误解，于是采用了“定名”、“定序”、“定距”和“定比”。翻译中的问题就交代这么些。下面再交代一下本书的翻译过程和译者间的分工。

这项不大的翻译工程是汪顺玉博士发起的。早在五六年前，汪博士就提议和我合作写一本社会统计学的入门教材。由于我“深陷”社会测量学的基础理论研究之中，一直没有精力考虑如何把他的这项美好提议变成现实。后来，他主动退了一步，提出，如果没有精力写，那就翻译一部。我接受了他的这个“让步”，而且附加了一个条件，那就是，由他自己负责选书。在广泛调研的基础上，汪博士从大量的英文版统计学著作中，选定了纳迪的这本《如何解读统计图表：研究报告阅读指南》，而且自愿负责初稿的组织翻译工作和出版事宜的联系工作。其后，他不仅承担了第 5 章的翻译，还对初稿做了初步统一。

初稿的翻译是这么分工的：第 1 章由孙星翻译，第 2 章由佟美娴翻译，第 3 章由肖阳田翻译，第 4 章由夏语翻译，第 5 章由汪顺玉翻译，术语表由罗伊琳完成，席仲恩负责全书的最后统稿和定稿工作。

重庆大学出版社的雷少波先生和林佳木女士为本书的顺利出版做了很多工作。他们不仅迅速解决了本书的版权转让问题,而且在本书的规范和编辑上倾注了不少心血。对于他们的工作和贡献,我谨代表本书的全体译者表示最衷心的感谢!

尽管我们翻译时做了多方面的努力,但由于学识所限,译文中的缺点错误在所难免,欢迎大家不吝指正。

“工欲善其事,必先利其器”。统计学是我们进行社会科学研究的“器”,也是我们解读社会科学定量研究报告的“器”,更是我们信息时代提取信息和充分利用信息的“器”,而《如何解读统计图表:研究报告阅读指南》正是我们“利器”之“器”。愿大家充分利用这本利器之器,早日为我们自己打磨出一具如意的统计学之器。

译　者

2009 年 3 月于重庆邮电大学

作者前言

欢迎阅读《如何解读统计图表:研究报告阅读指南》。

本书附有我们的《研究检索指南》。"指南"可以帮助您从数以百计的学术刊物、通俗杂志以及刊载研究数据的报纸上搜索并提取文章。您可以用"研究检索"上的搜索引擎,通过任何一台与国际互联网联通的计算机查找您所需要的文章。任何拥有《如何解读统计图表:研究报告阅读指南》的读者,都可以享受六个月免费使用"研究检索"的权利。"研究检索"包括了两大类本书所讨论的科学数据资源:一类是学术论文,另一类是大众出版物。

学术论文

EBSCO 的 ContentSelect(内容摘要)学术期刊数据库是一个硕大的档案室,储存了大量通过专家审稿的学术刊物的论文,所有论文按照学科分门别类。这些文章按照严格的科学途径,向我们提供了关于具体研究课题的专门知识和信息。您可以用文章标题、作者、主题,或者每篇文章的特定代码[AN]在"内容摘要"(ContentSelect)中搜索所需要的文章。本书选取的许多例子和练习,其发表的论文都可以通过"内容摘要"(ContentSelect)获得。

大众出版物

报纸和杂志是定期出版物(每日一期,每周一期或每月一期),它们的目的是为普通读者提供信息。"内容摘要"

(ContentSelect)中有一个“共同兴趣”主题数据库,包含诸如《新闻周刊》《基督教科学箴言报》以及《今日美国杂志》这类报纸杂志上所发表的文章。“研究检索”还包括一个“主题搜索”模块,收录了一年的《纽约时报》内容,供大家使用。

除此之外,“研究检索”还提供广泛的在线服务,包括:

- 启动研究工作
- 寻找、评价信息来源
- 引证信息来源
- 通过互联网做研究
- 创建个人“图书馆”
- 论文写作指导

点击 www. resarchnavigator. com,进入“研究检索”。使用《研究检索指南》首页提供的代码进行注册,根据屏幕上的提示建立你的免费使用。

致 谢

本书是我近 30 年来在皮兹尔学院为本科学生教授统计学以及量化研究方法的成果。皮兹尔学院是克莱蒙学院联盟中的一员。通过学习如何解读统计量的意义,学生开始明白如何更好地开展研究、分析数据。他们的耐心、好奇心以及坚持不懈的求学精神促成了本书。

本书的存在还要感谢我的编辑杰夫·莱萨尔和他的同事们,感谢他们对我的想法的不懈支持,以及他们在本书出版过程中的专业监督。

最后,这也是需要说明的最重要的一点:没有杰夫·切明,就没这本书,因为有那么多的问题他想知其所以然。现在本书也能解答您的同类问题了。

目录

导 论

在如今这个愈发数据化的世界里，读懂报纸和杂志文章中的数字图表，解读学术出版物里的统计结果的意义，理解统计软件输出的统计结果，都是极其重要的技能。对于很多人来说，解读数据如同学习一门外语，非得亲自实践、亲身体验真实的问题情景不可。

目 的

《如何解读统计图表：研究报告阅读指南》是为了协助读者解读大众读物和学术文章中的基本统计结果、表格和统计图而设计的。本书语言浅显易懂。那些希望能快速了解数据解读技巧的读者，以及正通过标准的研究方法或统计学教材学习统计学的读者，可将本书作为补充读物。本书既不是一本关于如何做统计分析的书，也不是一本关于如何使用统计软件的指南。但是，本书的确提供了大量的 SPSS 统计软件的输出结果和学术出版物中的实际统计结果，同时还包括了来自诸如民意调查和调研报告这类源自日常生活问题的统计图表。本书的目的就是使读者能够读懂这些图表，能够充分地理解这类图表。

虽然读者可以以任何顺序阅读本书,不过前几章为解读数据表格以及进行描述性统计分析提供了良好的铺垫。随后的几章介绍了一些更为高级的统计技术以及解读它们意义的方法。如果在阅读《如何解读统计图表:研究报告阅读指南》一书的同时再读一本统计学教材,书中的信息将得到强化,因为统计学教材可澄清许多概念。当然,也可以在互联网上搜寻到各式各样的在线统计学书籍、研究方法书籍和相关的统计学术语表。

内容编排

每章都按照以下顺序介绍一个统计量、统计表、统计图或一组统计数字:

a. 首先提出定义;

b. 介绍该统计量(统计表、统计图或统计数字)适用和不适用的条件或假定;

c. 通过三种方式理解它的用法:用大众出版物中的实例(如果可以收集到的话),用学术论文中的实际例子,用 SPSS 软件的统计结果输出;

d. 解读统计量、统计表或统计图的意义。

对每个统计量、统计表或统计图的解读又包括:

a. 首先找出变量,说明使用该统计量、统计表或统计图的原因;

b. 用文字叙述统计表或统计图中的内容;

c. 用日常生活语言作总结;

d. 说明或假设产生该统计结果的原因。

除此之外,每一章都包含了实例作为练习,以便读者边学边用,最终达到学以致用。练习都放在练习框内。每章都穿插了多个练习,以衡量各个阶段所取得的学习进展。

定义

理解《如何解读统计图表:研究报告阅读指南》中的材料,要求我们熟悉一些专业术语和概念。这些术语和概念通常在研究方法和统计学课程中学到。考虑到一些读者需要温习一下这方面术语和概念,而另一些读者尚未学习过研究方法和统计学,下面,我们首先介绍一些关键术语。

变量:变量是可以取不同值的可测概念。在研究的样本中,变量取不同的值。例如"宗教"这个变量就有几种可能的取值:天主教、佛教、犹太教、清教、伊斯兰教以及其他。变量测度行为、态度、观点,描述参与者的人口统计学特征(如年龄、种族/民族、性别)。

值:值表示一个概念的变异性。变量"政党"可取"共和党"、"民主党"、"保守党"、"工党"这样的值,或者其他任何一个适合样本情况的值。区别"变量"和"值"的一种简单方法是:调查的问题通常是变量,对问题的回答一般是值,请一定记住这一点。

自变量:对一个变量进行解释或预测时要用到自变量这个概念。一项调查可能会用"民族/种族"来预测或解释"投票行为"。这里的"民族/种族"就是自变量。在该例中,它在投票进行之前就已经存在了。

因变量:因变量是这样一种概念,这种概念的发生取决于其他变量。因变量包含着由自变量引起的结果或自变量所预测的值。在上例中,"投票行为"就取决于种族/民族,或者是由种族/民族引起的。

定类变量:定类变量指不具有数量或顺序意义的类别或特征。"所拥有的车种"这个变量可以取多种值,如"本田"、"雪弗莱"、"德国大众"等。这些名称之间并没有特定的顺序,也没有数学或数量属性。它们仅仅是用来测度"所拥有的车种"

这一变量的。为方便计算机分析,我们可用一个数字对每一种回答编码。不过,这个数字是任意的,它并不具有任何的数学属性。“本田”可以编码为1,3或12。各种回答之间并没有顺序,编码为4的回答并不是编码为2的回答的两倍。

定序变量:定序变量是指具有顺序意义的类别。因为具有顺序性,我们就可以用一个数字代表一个类别,以表示这种顺序性。例如,一个调查问题可能会问过去一个月你下馆子的频繁程度。这里“下馆子的频繁程度”是变量,它的取值或回答是“常常”、“有时”以及“从不”。我们可以用数字1代表类别“常常”,2代表“有时”,3代表“从不”。他们是根据频繁程度排序的,因此,你不能用2来代表“常常”,用3代表“有时”,用1代表“从不”。第一个数字和它所代表的类别之间的关系通常是任意的,但既然这么开始了,你就得有始有终,用数字标出各类别之间的顺序。因此,你可以选择用1表示“从不”,但同时就应该用2表示“有时”,用3表示“常常”。畅销书排行榜以及前49位CD销售排行榜就是两个定序变量的例子。

定距变量和定比变量:定距变量和定比变量具有数学和数量特性,其取值通常是由某种计量系统实际测得的区间大小。例如,用“英寸”表示的高度,用“磅”表示的重量,用“岁”表示的年龄。定距量数和定比量数之间的唯一区别是,定比量有绝对零点,这意味着在你的测量结果中不能有负值,即不能出现-50磅或-2岁这种情况。绝对零点的存在使得我们可以说一个10岁孩子的年龄是一个5岁孩子年龄的2倍(比率)。对于定距量来说,零点不过是一个真实的答案。但是,如同温度一样,你不能说20 ℃比10 ℃热2倍。定比测量是唯一具有加、减、乘、除属性的变量,因此,在一些更为高级的统计分析中备受人们的青睐。不过,对于所有的实用目的,可以用同样的方法处理定比变量和定距变量的测量结果。为简明起见,本书统统用定距/定比表示这两类量数。有些定序量数,如李克特(Likert)态度级标(从“强烈赞成”到“强烈反对”)的测量结果和两分判断(即两值变量,如真/假、男/女)结果,都似乎显示出

了等距关系,因此,也经常像处理定距/定比量数一样处理这类结果。

假设:假设是对两个或两个以上变量之间关系的陈述,这种陈述未经检验。假设来自于理论、直觉、观察或以往的研究。多数时候,我们用否定形式表述假设,例如,"对于资金处罚所持的态度与受教育程度之间没有关系"。

总体:总体是你要分析的单元或元素的全部。总体取决于所有可能元素的可获取性。例如,要对你家乡这个总体进行调查,就要求有一份你家乡全体人口的名单。总体的特征叫参数。

样本:我们更可能是对一个样本进行研究,而不是对总体。当总体的数量很大时更是这样。一个样本由组成总体的全部元素或单位的一个部分构成。在理想情况下,样本根据随机抽样原则(总体中的每一个因素都有同等机会被抽取)选定。对于这种随机选取的样本,我们就可以根据样本变量的取值去推断总体参数的值。对样本特征进行描述的那部分统计学称为描述统计,通过样本情况去对总体情况进行推断的那部分统计学称为推断统计。

概括性:概括性是指用样本统计量对总体参数进行精确推测和准确描述的能力。但是,如果样本不是随机选取的,而是根据非概率抽样技术选取的,例如方便性抽样、目的性抽样、配额抽样或滚雪球抽样,那么,你的结论就只能局限于该样本。

抽样误差:抽样误差是指样本统计量(即数据)与总体实际参数值之间的差。我们的目的是限制误差的大小,以便对所研究的总体中各种关系进行更准确的预测和解释。大多数的民意调查在宣布最后结果时都会报告相应的抽样误差,如"正、负4.5个百分点"。

显著性水平:显著性水平即 p 值,它表示凭机遇或偶然性获得某特定统计结果的概率。所有的统计结果都由两个部分组成:通过数学公式计算得出的值,以及凭偶然性获得这个值的可能性。这要由各种不同的概率模型确定,如正态钟形曲

线。例如,一个卡方值可能是 3.45,相应的显著性水平是 0.024。这个结果用$\chi^2 = 3.45, p < 0.05$ 来表示。它告诉我们,仅凭偶然性得到3.45 这个卡方值的概率小于 5%。大多数社科研究中都规定 0.05 是最低标准,其意义为:根据从样本上收集的数据对总体情况作推断,可以接受的错误率最高是 5%。如果认为这个错误率过高,可以规定其他更低的标准,如 0.01 或0.001。如果达到了事先规定的显著性水平,我们就可以宣布,该研究结果具有统计显著性,并拒绝零假设。据此,我们便可以得出结论,所研究的两个变量之间的确存在某种关系。所设定的显著性水平也就是发生第一类错误的概率,即,在两个变量之间没有关系时,你却拒绝接受这个事实。显著性水平是你宣布你的发现成立但实际上并不成立的百分比。

书尾的参考文献包括了其他一些关于统计学和研究方法的著作,供大家进一步参考之用。

统计分析决策树型图

解读统计信息的关键是正确理解测量结果的水平,选择适当的统计分析技术。图 1 是一张“统计分析决策树型图”,它能帮助我们逐步解读统计分析的过程,在各个关节点上作出正确的选择。

确定哪个是你提出的假设或研究问题中的自变量，哪一个是因变量。

确定每个变量的测量水平（定类、定序、定距/定比）。
请牢记，二分测量结果可以作为任何水平进行处理。许多看似等距的定序量数也可以当作定距/定比量进行分析。

进行频数和其他合适的描述统计分析（如众数、中位数、均值），以检查样本的每一个变量是否的确有变化，而不是恒定不变。

选择适当的统计方法，分析自变量和因变量之间的关系。

比较同一个因变量两个类别间的均值（定距/定比）：用t检验。
比较同一个因变量三个或三个以上类别间的均值：用方差分析。

检验定类变量及/或定序变量间的关系：用卡方检验。

检验两个定距/定比变量之间的关系：用皮尔逊积矩相关。
如果有两个或两个以上的自变量：用复相关和线性回归。

如果凭偶然性获得该统计值的概率低于0.05（$p<0.05$），那么就拒绝零假设，宣布两变量之间存在显著关系。

对于皮尔逊积矩相关，还要评估关系的强弱：0~0.25为低相关，0.25~0.60为中相关，0.60~1.0为高相关。
正、负（+、-）号还告诉我们该相关是正相关（同时增加或同时减小）还是负相关（随着一个变量的增加另一个变量减小）。

图1 （摘自《进行调查研究：定量方法指南》，彼德·M.纳尔蒂，阿里恩培根出版社，2003年）
（from *Doing Survey Research: A Guide to Quantitatice Methods*, Peter M. Nardi, Allyn & Bacon, 2003）

第1章　描述数据

在调查问卷返回到研究者手中并开始下一步的数据分析之前，研究者通常会先评估一下问卷中各个项目的回答情况，确定它们是否的确是变量，也就是确定每一个项目都具有足够的变异性以供进一步的数据分析。举个例子，假如一项民意调查的结果显示有95%的调查对象表示“强烈赞成”，或者，假如回答该调查的人数中有90%的人为女性。这种情况通常不需要进行分析，除非样本容量非常巨大。一个项目要对随后的数据分析有用，回答者对这个项目的回答就必须有一定的变异。

还有，除了对所要研究的其他基本情况进行描述外，对完成调查或问卷者的人口统计学属性也要作较详尽的描述，这是撰写研究结果报告的一个重要步骤。我们可以通过图形、描述性统计量以及频数表来实现我们的目的。本章将引导你正确解读以各种形式展示的一元数据，即一次只呈现一个变量的数据。

定　义

■ **趋中量**:趋中量有均值、中位数和众数三个,它们总结性地描述了变量取值分布的集中位置。

■ **均值**:均值描述了一个分布的数值中心。将分布中的所有值之和除以值的个数就得到该分布的均值。

■ **中位数**:如同高速公路中心的分界一样,中位数是一组按顺序排列的数据分布的中界。中位数是把全部数据一分为二,大于它的放在它的一边,小于它的放在它的另一边,按照大小顺序,以中位数为中心向两边分布。

■ **众数**:众数是一个在变量分布中出现次数最多的那个值。千万不要把众数和"多数"相混淆。多数表示 50% 以上的接受调查者取这个值,多数一定是众数。但出现频次最高的值可以低于 50%,却仍然不失为众数。

■ **标准差**:标准差告诉我们一个分布的分散情况。标准差为我们提供了一个衡量数据分布分散程度或差异程度的数学量。标准差有点像数据分布中所有数据相对于均值的差的平均值。

■ **百分位数**:百分位数告诉我们调查结果中有百分之几的值位于某一个特定值之上,又有百分之几的值处于这个特定值之下。例如,某个值的百分位数是 50,这意味着有 50% 的回答处在该值之上,也有 50% 的值处于该值之下。

■ **频数表**:频数表或频数分布都表示调查对象对某个问题(即变量)作出的各种回答(即变量取值)的次数情况。表中包括频数和频率。每个值的频数就是这个值在原始数据中出现的绝对次数,每个值的频率就是这个值的频数占总数的百分比(有时称之为有效百分数),或者这个值的频数占总调查人数(包括没有回答该问题的人)的百分比。

■ **条形图**:条形图又叫柱形图、直条图,是呈现变量分布情

况的一种直观方法。通常,变量沿 x 轴(即水平轴)分布,变量不同取值的频数(用百分数或原始出现次数表示)沿 y 轴(即纵轴)分布。顾名思义,条形图就是用条形的高度表示一个值与其他值出现次数的多少。有时条形图是水平放置的,这种情况下,沿 y 轴分布的是变量的类别或取值,沿 x 轴分布的就成了频次或频率了。于是,条形的长度(并非高度)表示变量取值发生的频次/率。对各条形图作比较时,一定要保证频次/率轴(通常是 y 轴)单位的相当或一致。

- **直方图**:直方图与条形图类似,不同之处是,构成直方图的所有条形都紧紧地挨在一起,以显示所示量是连续的。条形的宽度和高度一起告诉我们某个特定区间上的应答人数。
- **频数多边图**:频数多边图又叫折线图,是通过连接直方图中每个条形(或区间)的中点构成的一条折线。社会科学研究中常用的频率图是正态曲线图或倒钟形曲线图。通常,我们用多边图来描述变量随时间发生的变化。
- **饼图**:饼图又叫圆形图或扇形图。饼图是通过把一个圆分割成几个扇块,用扇块面积的大小表示变量不同取值的发生频次/率,是一种描绘变量取值情况的直观方法。

假　定

各种统计图和描述性统计量适合于展示或描述一元变量的信息,即一个图一次只展示一个变量的情况。应用中到底要选择哪一种方法,通常取决于变量测量结果的水平。

- **定类量**:使用众数、饼图、条形图。
- **定序量**:使用中位数、众数、饼图、条形图。
- **定距/定比量**:使用均值、中位数、众数、标准差、百分位数、直方图(对于连续变量)、条形图(对于离散变量)、频数多边图、曲线图。

取自大众出版物的真实数据

饼图

每年,美国教育部都会通过国家教育统计中心发布美国公共教育各方面的消息。报道中的各种数据表通常都伴有相应的统计图,使数据直观化。在继续向下阅读之前,请仔细看一看图 1.1 的圆饼,看看你是否能够解读这张饼图。

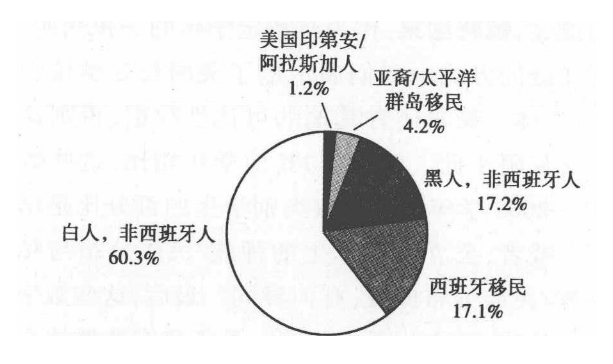


图 1.1 2001—2002 学年各种族/民族学生占公立学校小学生和中学生的百分比

来源:美国教育部国家教育统计中心共核数据①处(CCD),"国家非财政年度公立小学/中学教育调查",2001—2002 年

Source: U. S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD), "State Nonfiscal Survey of Public Elementary/Secondary Education,"2001-02.

找出变量:饼图适合于一次表示一个变量。在这个案例中,饼图表示 2001—2002 学年公立学校小学生和中学生的种

① "共核数据"是英语 Common Core Data 的翻译。一项大规模的社会调查通常要分解为若干个子项目进行。为了提高调查结果的利用效率,通常把研究各个子项目时都可能用到的数据集中到一个数据模块,这个数据模块就是"共核数据"模块。通常,"共核数据"是和"专项数据"相对而言的。

族/民族的分布情况。一份优质的统计图应该在标题中提供足够的总结性信息,并在来源处注明数据的具体出处。种族/民族是一个定类水平的量,因为该变量的各个类别间并没有固有的先后顺序。

解读图表:即使不看百分比,我们也能一眼看出,饼图中最大的扇块表示非西班牙裔白人学生,其次为非西班牙裔黑人学生,然后是西班牙裔学生。

饼图适合于表示数量信息(在此例中为百分比),不但简单快捷,而且直观。许多饼图并不注明百分比。如果要给出各类别所占的确切百分比,最好用频数表。

得出结论,解释结果:作为某特定学年的一张快照,这幅饼图以方便快捷的方式向我们展示出了美国公立学校中的最大种族/民族群体。除非还有其他的可比性数据,否则这幅图对我们的意义仅限于此。例如,与其他学年相比,这些数字是否表示 2001—2002 学年每个种族类别学生的百分比是增加了还是减少了? 或者,公立学校学生的种族/民族分布与私立学校学生的种族/民族分布比较,有何异同? 最后,这些数字是否与总体人口的种族/民族分布成比例? 是不是有些群体在公立学校中的人数过多或者不足?

条形图

对频率分布情况还可以用条形图来进行描绘。当有几层数据需要同时进行描绘和对比时,条形图尤其有用。图 1.2 是一幅注明了频率数的条形图。

找出变量:这个图显示了在进行这次调查时,年龄为 18 ~ 24岁的高中毕业生在各种教育程度上的百分比。该图比较了在家只说英语和在家还说其他语言这两个群体。因此,变量为“教育程度”和“家庭语言”。这里,年龄不是变量,是常量,因为这个样本中的每个人都处在同一年龄范围内。此外,“高中毕业”也不是变量,因为每个人至少都高中毕业。

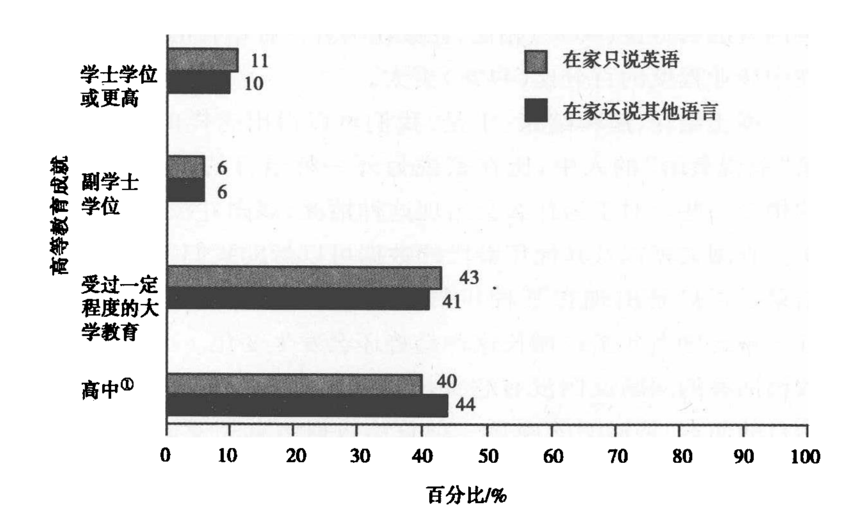


图 1.2 1999 年 18~24 岁青年教育程度分布

注:①高中毕业生包括通过同等学力测试(如 GED 考试)的学生。

由于四舍五入原因,各教育程度百分比的总和不一定正好是 100。

来源:美国商业部人口普查局当前人口调查组(CPS),1999 年 10 月。

Source: U. S. Department of Commerce, Bureau of the Census, Current Population Survey(CPS), October 1999.

解读图表:我们从两个变量(教育程度和家庭语言)中的任意一个入手。即使不看百分比值,条形的长度也告诉我们,大多数人都获得了高中以上文凭。表示"高中"的条形的长度与"受过一定程度的大学教育"的大体相当,表明现在正读大学者或已经受过一定程度的大学教育者的百分比与只有高中毕业的人的百分比基本相等。但要注意的是,有些人已经获得了副学士学位(相当于我国的专科毕业)或学士学位。假定抽样对象为 40 多岁的中年人,那么这份调查的条形图又会是怎样的呢?

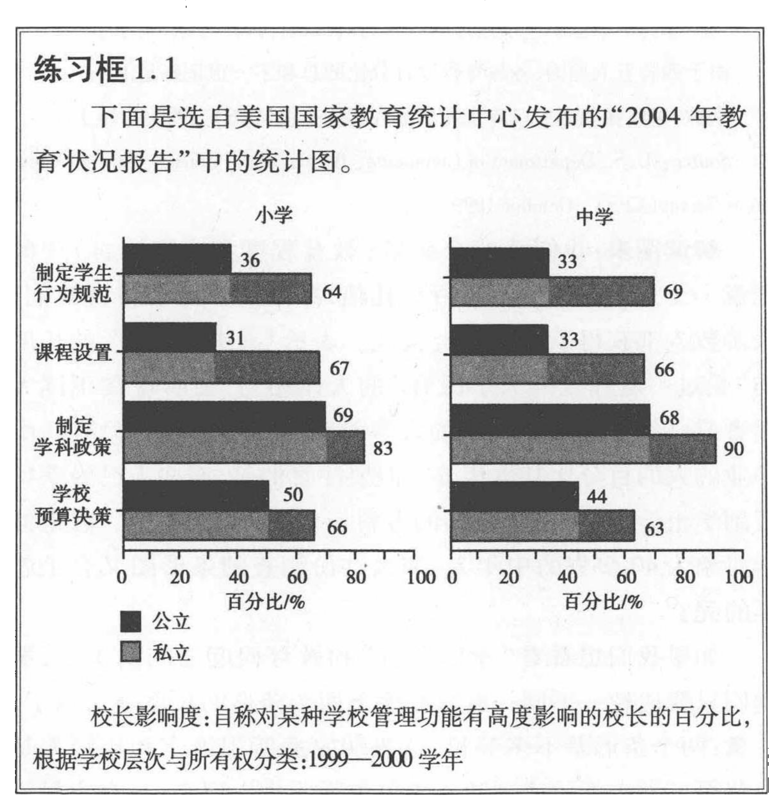
如果我们想看看"家庭语言"和教育程度之间的关系,那我们只要比较一下同一教育程度上两个条形的长度就可以了。注意:两个条形差不多等长,这表明家庭所用语言和年轻人是否获得了学士学位或副学士学位关系不大。但是,与在家只说

英语者的百分比(40%)相比,在家说另外一种语言的人停留在高中毕业程度的百分比(44%)更大。

得出结论,解释结果:于是,我们可以得出这样的结论:在家“只说英语”的人中,比在家说另外一种语言的人上过大学的稍多一些。对于为什么会出现这种情况,该图并没有提供答案。查阅文献以及其他年龄段的数据可以帮助我们判断,这个结果是否只是出现在某种年龄段(只适用于 18 ~ 24 岁的人群),或许随着年龄的增长这种趋势还会发生变化。可见,这组数据回答的问题反倒没有它提出的新问题多。不过就其描述的目的而言,条形图的确是一种总结数据信息的便捷、直观方式。还要注意的是,这个图并没有提供任何显著性检验信息,因此,我们还不能得出一个统计学结论:在家只说英语的人比在家说其他语言的人接受了更多的教育。

练习框 1.1

下面是选自美国国家教育统计中心发布的“2004 年教育状况报告”中的统计图。



校长影响度:自称对某种学校管理功能有高度影响的校长的百分比,根据学校层次与所有权分类:1999—2000 学年

问 题

1. 该图表示的变量是什么？它们是何种测量水平(定类、定序、定距/定比)上的变量？
2. 为什么条形图适合本调查？
3. 用文字描述该图所表示的内容。哪些条形最长？哪些条形最短？例如,左图中标有 66 的条形与它上方标有 50 的条形相比,告诉了我们什么？
4. 用日常语言总结这些结果。小学校长与中学校长之间存在区别吗？在各层次的学校中,公立学校校长与私立学校校长的最大区别在哪些方面？
5. 对这些结果可能做出的合理解释是什么？有什么理论或以前的研究结果可以指导你解读这些发现？

折线图

用折线图表示随时间变化的数据,通常是一种较好的选择。图 1.3 为我们直观地呈现了美国合法移民中被接受为永久性居民的情况。该统计图源自《1998 年统计年鉴》,选 2004 年的一份题为《少数民族语言使用者及其教育和劳动市场指标——近期趋势》的报告。该报告由美国国家教育统计中心发布。

找出变量:合法移民人数是一个定距/定比变量,年度也是一个定距/定比变量。由于有时间维度,折线图最能揭示这组数据的意义,因为,折线图能够直观地表示移民人数随年度的连续变化趋势。

解读图表:从 1988 年到 1991 年,移民人数连续四年急剧增长;1992 年,移民人数大幅下降。尽管还没有下降到 1988 年以前的水平,但随后,除 1996 年有微弱增长外,移民人数持续走低。注意,通过这幅图要确定每年的确切人数是不可能的。频数表可以提供具体的人数,把确切人数标注到折线图上也可以。统计图旨在直观地呈现数据的总体情况,使读者一目了

然，而不在于给出具体的频数/率信息。

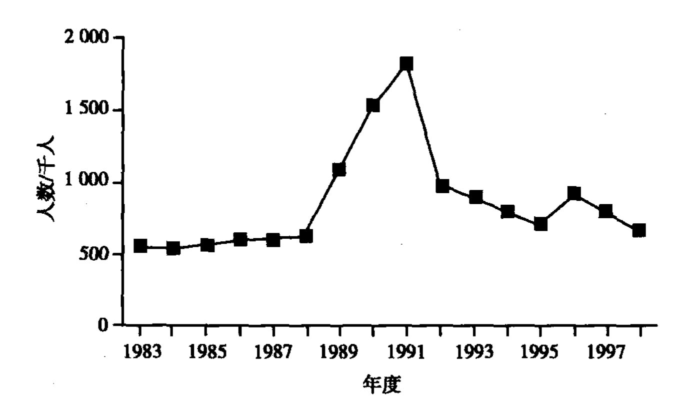


图 1.3　1983—1998 年美国移民人数变化趋势

注：图中描绘了 100 000 移民的情况。“移民”指被批准在美国合法永久居住的非原籍美国人。

来源：美国司法部移民规划服务，1998 年统计年鉴，华盛顿特区：美国政府印刷办公室。

Source: U.S. Department of Justice, Immigration and Naturalization Service. 1998 Statistical Yearbook. Washington, DC: U.S. Government Printing Office.

得出结论，解释结果：或许，应该把解释集中在移民法的变动、选举年的活动或其他相关的政治事件上。这幅统计图并没有提供任何解释说明，因为它的目的主要是描述。该折线图向我们直观地展现了 16 年间美国合法移民人数的变化情况，使我们一目了然，迅速获得这一变化信息。

频数/率表

频数/率表能方便地总结和比较两个不同时间点上的趋势。表 1.1 展现的是美国在家接受教育的学生人数变化情况。该信息源自美国国家教育统计中心发布的一份政府报告。

表 1.1 幼儿园至十二年级,年龄在 5 ~ 17 岁之间在家学习学生的数量和百分比分布情况(根据 1999 年和 2003 年入学注册状况)

入学注册状况	在家学习的学生			
	1999 年		2003 年	
	数 量	百分数	数 量	百分数
总数	1 003 000	100.0	1 294 000	100.0
完全在家学习	697 000	69.5	898 000	82.0
部分时间在校学习	153 000	15.3	198 000	18.0
每周在校少于 9 小时	107 000	10.7	137 000	12.5
每周在校 9 ~ 25 小时	46 000	4.5	61 000	5.6

注:不包括公立学校或私立学校每周在校时间超过 25 小时的学生,也不包括因病暂时在家学习的学生。由于四舍五入的原因,各项之和可能不正好等于总数。标准误差可在 http://nces.ed.gov/pubsearch/pubinfo.asp? pubid-2004115 上获得。1999 年全国共有 50 188 000 名 5 ~ 17 岁的幼儿园到十二年级的学生,2003 年共有 50 707 000名。

来源:美国教育部国家教育统计中心 1999 年全国家庭教育调查(NHES)之家长调查,2003 年全国家庭教育调查之家长与家庭教育参与度调查。

Source: U.S. Department of Education, National Center for Education Statistics, Parent Survey of the 1999 National Household Education Surveys Program (NHES); Parent and Family Involvement in Education Survey of the 2003 NHES.

找出变量:变量为幼儿园到十二年级学生(5 ~ 17 岁)的"入学注册状况",变量的取值为"完全在家学习"和"部分时间在学校学习"。"部分时间在学校学习"又细分为"每周在校少于 9 小时"和"每周在校 9 ~ 25 小时"。该表对两个不同年度(1999 年和 2003 年)的数据进行了比较。根据表注,该表并没有包括每周在校时间超过 25 小时的学生和因病暂时在家学习的学生。

"入学注册状况"是一个定类变量,很适合用频数/率表呈现调查结果。通常,原始频数、百分比以及总数都要列在表中。不过总数放在了数据部分的第一行,而不是最后一行。有时,

总数也放在数据部分的最后一行。

解读图表:就数量而言,四年时间里在家学习的学生出现了大幅度的增长。2003 年,百万以上的学生选择了在家学习,而 1999 年的总数只有 850 000 人。这些人中,82% 的人完全在家学习,而未进任何公立或私立学校,18% 的学生每周到学校学习一段时间。这两个百分比在 1999 年和 2003 年都一样。虽然在数量上出现了急剧的增长,但就百分比而言,完全在家学习的学生在 1999 年和 2003 年并没有变化。每星期到学校学习的那 18% 的人,2003 年分化为 12.5% 的人每周在校时间低于 9 小时,5.6% 的人每周在校时间为 9 ~ 25 小时。这些百分比几乎与 1999 年的百分比相同。

得出结论,解释结果:在美国,虽然在家学习的学生的数量出现了增长,但完全在家学习而不参加学校教育的人的比例并没有发生变化。虽然这些学生中有的完全在家里学习,不过,还是有小部分学生同时参加了学校教育。

这张表并没有提供总人数上升的原因。部分原因是总体人口的增长,但即使考虑到总人口的增长,在家学习的学生人数还是上升了一些。表中未列出的此次调查的其他数据表明,或许是对学校环境的担忧,再加上想为孩子提供宗教和道德教育的愿望,促使越来越多的父母选择把孩子留在家里自己进行教育。

其他频数表与统计量

还有很多种展示频率信息和描述统计结果的方法。表 1.2a和表 1.2b 展示了从国家教育统计中心共同核心数据处获取的有关数据。这一次,我们关注的是“2000—2001 学年美国最大的 100 所公立小学和公立中学的特征”(表 1.2a)。

表 1.2a

	2000—2001 学年		
	全部学区总人数	最大的 100 个学区总人数	最大的 100 个学区人数占全国总人数比重
学生	48 067 834	11 050 902	23.0
教师(全职)	3 002 947	641 333	21.4
学校	95 366	15 615	16.4

表 1.2b 2000—2001 学年美国地方 100 个最大学区公立小学和公立中学的生师比的中位数

全部学校	小学	初中	高中
16.7	16.9	16.6	17.4

找出变量:这个例子中的变量是全部学区以及这 100 个最大学区中、小学生人数,教师人数和全部学校数量。由于这些数据都是实际的计数结果,所以变量为定距/定比量。事实上,表 1.2b 计算出了生师比,表明了各类公立学校中,平均每个老师教多少个学生。

解读图表:表 1.2a 列出了学生、教师和学校的原始数据,以及全美国 100 个最大学区中的学校数、全职教师数和学生数占全国总数的百分比。例如,在 2000—2001 学年,美国国内全部学区共有 95 366 所学校,100 个最大学区内共有学校 15 615 所。15 615 除以 95 366 就是 16.4%。

表 1.2b 列出了全部学区学校的学生教师比,并根据学校层次进一步细分。表中报道的是中位数比,即处在第 50 百分位上的生师比。例如,在美国,2000—2001 学年有一半的公立高中,平均每个教师只有不到 17.4 个学生,而另一半公立高中的生师比要高于 17.4。这个比值是通过把每所学校的总学生人数除以总教师人数得到的。然后根据比值大小,从高到低对学校进行排序。中位数比是这样一个点,在它之上有一半的学校,在它之下有另一半学校。本例中,这个中位数点就是 17.4。

比值显然是定比量数,所以,我们可以计算比值的均值。但由于极端数据扭曲了均值,因此,负责报告的机构认为用中位数更好。

得出结论,解释结果:我们可以看出,在最大的前100个学区学校就读的学生占全国总在校生的23%,但他们所就读的学校数只占全国学校总数的16.4%,教师人数只占全国教师总人数的21.4%。这或许向我们揭示,在最大学区的学校中,班级较大,学生较拥挤,也就是说,生师比较大。

虽然表1.2b提供的原始报告中分学区报道了这些数据,但是,表1.2b中并没有提供这些数据。不过,表1.2b的确告诉我们,高中的生师比大于小学、初中。这个结论很有意义,因为在美国的各个区中,中学数量通常要比小学数量少。

练习框1.2

大多数研究都用一张表格汇总受访者的关键人口统计学信息,以简明扼要地描述样本。这是一张由MDRC(一个大型研究机构)作的福利改革纵向研究报告中的统计表。在这次研究中,研究者用结构化访谈方式收集调查数据,并对福利受益人做了定性半结构化民族志访谈。表中描述了两个样本的特征。

特　征	调查的样本	民族志样本
平均基础年龄/岁	30.2	35.8
种族或民族群		
非裔/%	69.0	45.2
白人/%	2.1	0.0
西班牙裔/%	28.3	54.8
其他/%	0.7	0.0
基线儿童的平均数量	2.7	3.0
基线最小儿童的平均年龄/岁	4.9	6.0

续表

特　　征	调查的样本	民族志样本
基线具有高中文凭或 GED 者/%	48.9	42.9
1994 年 4 月—1995 年 3 月获得失业保险收入者/%	39.9	*n/a*
样本容量	581	42

来源:城镇变化调查与民族志访谈调查结果汇总。

Source: Calculations from the Urban Change Respondent Survey and ethnographic.

注:调查样本的起始期为 1995 年 5 月,民族志样本的起始期为 1997 年 10 月。

对调查样本的最后访谈在 2001 年 3 月—10 月间进行,对民族志样本的最后访谈在 2001 年的 10 月进行。

由于四舍五入缘故,总量与分量之和可能有微小差别。

来源:Brock, et al. (2004)

问　题

1. 在表格中哪些是变量? 它们是哪个水平的量?
2. 为什么均值和百分比适合于这些调查结果?
3. 用文字描述这张表告诉了我们些什么,两个样本有哪些相同点和不同点。
4. 用日常语言表述这项研究的结论。

取自学术文章的真实数据

学术文章和学术著作在报告定量研究结果时,大多依靠复杂的统计分析。不过,文字描述或一张图表通常能提供一些关于样本特征、变量分布或其他一些单变量数据的初步分析结果的信息。

频数表

表 1.3 呈现的是一项父母对孩子性教育态度的调查结果,调查是在希腊进行的。研究的目的是为了了解父母对家庭在孩子性发展和性教育中的角色持什么样的态度。

表 1.3 父母对"哪些问题对于你孩子的性发展和性教育影响最大?"问题选项的选择数量(n)和百分比(%)

回答组	n(252)	百分比
媒体影响	74	16
缺少合适信息	61	13
对身心健康的危险	40	8
学校缺少性教育	38	8
负面社会影响(毒品、暴力、犯罪、变态等)	36	8
如何变成成熟的男人或女人	31	7
友谊	29	6
如何谈论性	24	5
第一次性经历	22	5
缺少道德价值	21	4
其他	98	20
应答项次总计	474	100

来源:Kakavoulis(2001:171) [研究检索索引号:4480158]

找出变量:这个问题要求父母指出他们认为自己孩子在性成长和性教育中所面临的问题,该问题共有 11 个备择选项,其中 10 项标明具体类别,1 项标为"其他"。因为各个选项间并没有固有的顺序,该问题是定类量,适合用频数或频率表汇总统计结果。因为有必要表示各选项的确切百分数,用统计表比用统计图效果更好。

解读图表:注意,表中告诉我们,共有 252 人(n)回答了这个问题,总计 474 应答项次。很清楚,人们可以选择一个以上他们担心对自己孩子的性成长和性教育有影响的选项。其中一种展示调查结果的方法是,有百分之几的应答者选择了某个选项。关于这个问题,252 个应答者中,74 人选择了"媒体影

响”选项，占应答者总数的 74/252 =0.294 或 29.4%。不过，该表格并没有这么做，而是报道了选项的被选择次数占总应答项次的百分比，即 474 个应答项次中，有 74 项次选择了第一个选项，占 74/474 =0.156 或 15.6%（表中四舍五入为 16%），换句话说，在应答者的总应答项次中，差不多有 16% 的应答集中在媒体的影响上。如果选择第一种统计方法的话，结果就会是：29.4% 的应答者认为大众媒体对他们孩子性成长的影响最严重。结果的顺序仍然是一样的，即这两种呈现方法的结果都是：大众媒体的影响是父母最担忧的因素。

如果对于一个问题的各种选择是互相排斥的，那就意味着只能且必须选择备择项中的一项，而不是“所有适合的项”。这种情况下，无论是根据应答总项次数还是应答者总人数来计算以上的百分比，结果都是一样的。因为在这种情况下，一个应答者和一应答项次完全对应，即一个应答者就是一应答项次。本例中，由于每个应答者可以选择一个以上的选项，因此，应答者总人数（$n = 252$）就不同于总应答项次数（总计 = 474）。表的标题也告诉我们，百分比是每个选项占父母回答总项次的百分比。

得出结论，解释结果：除了担忧自己的孩子会被不良信息误导以外，希腊父母似乎很担心媒体对孩子性教育所产生的影响。他们对其他备择项的关切比较均衡。

为什么父母认为孩子接触到的媒体信息和其他不良信息至关重要呢？该表并没有作任何解释。不过，有了这些结果，研究人员就可以利用它们，把它们和来自其他国家的数据作一比较，看其中是否蕴涵着文化异同。你所生活的社区的父母会给出很不相同的答案吗？

描述性统计量

表 1.4 是一个研究项目中用来测量工作对家庭所产生的溢

入效应①的描述性统计量。工作对家庭的溢入效应,就是多大量的工作会对家庭生活产生影响,会导致家庭问题。该项目研究了156对双薪夫妇的情况,具体关注点是工作对家庭凝聚力的影响。无需知道调查项目的确切措辞或数据是如何收集的,表1.4也为我们提供了关于各项得分的重要描述性信息,告诉我们这些分数的变异是否足够大,以便进行进一步的数据分析。

表1.4 分析所涉及变量的描述统计

变 量	男			女		
	均值	标准差	人数	均值	标准差	人数
因变量						
家庭凝聚力	27.29	4.25	150	27.75	4.56	150
自变量						
情感满意度	19.56	3.89	144	20.27	3.16	143
家庭满意度	2.52	0.84	153	2.77	0.66	146
工作对家庭的溢入	13.75	4.07	153	13.93	3.80	150
工作满意度	16.66	3.24	153	16.81	3.60	150
家庭和睦受益度	4.95	1.49	139	5.01	1.59	130
工作弹性	20.95	5.33	155	21.27	5.32	152
周均工作时间	46.23	13.51	156	36.20	13.35	155
孩子数量	1.44	1.46	156	1.56	1.59	156

注:量表中的缺失数据以应答者为单位删除。

分数越高,满意度越高。

来源:[研究检索索引号:8768458]

Stevens, Kiger, and Riley (2002)

找出变量:表中明确标出了因变量和自变量。根据研究报告的全文,"家庭凝聚力"是用一个由9个项目构成的量表测量的。量表为定距/定比水平,得分越高,家庭凝聚力越强。"情感满意度"是一个由6个项目构成的量表,集中点是家庭工

① "溢入效应"指的是工作对家庭生活带来的影响。站在家庭的角度看,这种影响是"溢入效应";站在工作的角度看,这种影响是"溢出效应"。如果"溢入效应"改善了家庭生活,那么,这种"溢入效应"叫"正溢入效应";反之,如果"溢入效应"妨碍了家庭生活,那么,这种"溢入效应"叫"负溢入效应"。

作的情感维度,特别是与孩子之间的情感。“家务满意度”量表只有 1 个项目,问的是他们对与家庭其他成员分担家务的满意程度。“工作对家庭的溢出”量表共有 6 个项目,得分越高,表明工作对家庭产生了较高的负溢入效应。“工作满意度”量表也是由 6 个项目构成。“家庭和睦受益度”只有 1 个项目,7 分表示对工作带来的益处(养育儿女、照料长辈等)非常满意。“工作弹性”是用一个由 8 个项目构成的量表测量的。

所有因变量都是定距/定比量。文章并未提供量表的项目样本,不过,感兴趣的读者可以直接从作者那里获得全部量表。

解读图表:解读表 1.4 的一种方法是,首先看一看每一个维度上的测量结果是不是具有足够的变异性,以便能做进一步的数据分析。标准差(SD)是一个重要指标,表明全部应答者(*N*)在各量表上的得分围绕其平均分(M)分布的情况。如果任何一个变量的标准差为零,这表明在该变量上不存在变异性,这些项目不能使用。脚注中的“缺失数据以应答者为单位删除”告诉我们,如果一个应答者(案例)没有对多项目量表中的一个项目作出回答,那么计算时,整个案例就会从该量表中删除。

另一种阅读该统计表的方法是,比较男性应答者和女性应答者的得分情况,看一看应答者的性别是不是和分数有关。似乎男女只有在每周平均工作时间上有差别,女性的平均工作时间要比男性的少。

得出结论,解释结果:每一个量表的描述统计结果都显示,本次调查中的男女在这些变量上都存在一定的变异性,因此,这些变量可以进行数据分析。除了每星期的工作时间以外,男女似乎都报告了差不多相同水平的满意度,无论是对于工作对家庭溢入的影响,还是家庭凝聚力。原文对这些自变量如何解释家庭凝聚力这一因变量的变异作了较详细的讨论。

练习框 1.3

有一项研究,目的是确定 72 名有超自然事物信仰倾向的大学生的人格特质。按照常规,该论文首先报告了样本的描述性统计结果。以下是这些学生在超自然物信仰量表上的得分情况,分高分组和低分组。

高分组和低分组超自然物信仰者
在普通问卷上得分的描述性统计结果($N=72$)

	低分信仰者	高分信仰者
性别	女 20	女 23
	男 16	男 13
自报 GPA(中位数)	3.00	2.90
您有宗教信仰吗?(中位数)	1.00(是)	2.00(差不多)
您每月参加几次礼拜?(中位数)	2~3 次/月	1 月/次
您相信超自然现象吗?(是)	33.3%	94.4%
您有同样信仰超自然现象的朋友吗?(有)	27.8%	100%
您阅读/观看关于超自然现象主题的材料或电视节目吗?(是)	41.7%	66.7%
您每星期花几小时阅读关于超自然主题的材料或观看相关的电视节目(15 小时或更多)?	50%	72.2%
在成长过程中你有过幻想中的朋友吗?	19.4%	41.7%

来源:Auton, Pope, and Seeger (2005)　[研究检索索引号: 11763486]

问　题

1. 有哪些变量?它们的取值是多少?每个变量的测量水平是什么?
2. 中位数告诉我们什么?例如,高分组的中位数 2.90 是什么意思?
3. 用文字描述该统计表向我们展示了什么。
4. 你能否得出结论说,高分信仰者和低分信仰者之间存在或不存在差别?

SPSS 结果输出

在 SPSS 软件包中，有多种方法可以展示一个变量的数据。统计图、频数/率表以及各种描述统计量，这些都是理解研究中变量分布情况非常重要的工具。

频数/率表

数据分析中的首要步骤是，看一看每一个项目的反应结果是不是有足够大的变异范围。下面，我们要考察一下表 1.5。这是一份《普通社会调查》(GSS)问卷上的一个调查项目，问调查对象有多少个孩子。

表 1.5　孩子数量

		频数	百分数	有效百分数	累积百分数
有效	0	414	27.6	27.7	27.7
	1	242	16.1	16.2	43.9
	2	398	26.5	26.6	70.5
	3	226	15.1	15.1	85.6
	4	115	7.7	7.7	93.3
	5	58	3.9	3.9	97.2
	6	14	0.9	0.9	98.1
	7	7	0.5	0.5	98.6
	8 或更多	21	1.4	1.4	100.0
	总计	1495	99.7	100.0	
缺失	NA	5	0.3		
总计		1 500	100.0		

来源：General Social Survey (GSS)，1993

找出变量:“孩子数量”是一个离散性的定距/定比量,尽管在严格意义上讲,“8 或更多”使得我们不能获得有 7 个以上孩子家庭中确切的孩子数量,不过,因为只有极少数人(1 500 人中只有 21 人)养育那么多的孩子,所以,它不可能扭曲按照定距/定比量所作的任何统计或分析。

解读图表:左边一列中的数字代表问题(变量)的有效回答(值),该列还包括回答问题人数的总和及没有回答问题的人数总和(NA 代表“没有回答”或“不适合”),最后是总计。第二列报告了每一个值的实际原始个数或频数,也报告了回答者的总人数、未答者的人数以及调查对象的总人数。

第二列的信息(标为“频数”)告诉我们共有 1 500 人(“总计人数”)参加了这次调查,但有 5 人(缺失)对该问题没有作出回答。在回答了调查问题的 1 495 人(总计人数)中,414 人没有孩子,242 人有一个孩子,等等。第三列(百分数)计算了每一种回答的人数占总人数的百分比,即 1 500 个调查对象中有百分之几的人作出这种回答。例如,有 5 人没有回答这个问题,所以 5 除以 1 500 等于 0.003 或 0.3%。1 500 个调查对象中,398 人有两个孩子,于是 398/1 500 =0.265 或 26.5%。

有时,这类信息很重要,但是,通常想要的是占真正回答问题人数的百分比。如果没有回答的人数很多,那么“百分数”一栏中的百分比就会歪曲事实。因此,SPSS 软件在计算“百分数”的同时,还计算“有效百分数”,即回答人数占实际回答总人数的“百分比”。例如,有 398 人回答说他们有两个孩子,398 除以 1 495 等于 0.266 或 26.6%。需要注意的是,当只有很少几个人没有回答时,百分数才与有效百分数十分接近。目的决定一切。有时,根据实际回答问题人数计算的百分数(“有效百分数”)很重要,有时候你却想知道基于样本人数计算出来的百分数,其中也包括那些没有回答问题的人。当“没有回答”本身被当作一种回答时,通常会出现这种情况。例如,关于民族/种族问题的调查结果,也许就需要标明拒绝回答者的百分比。

最后一列是“累积百分数”,表中是根据有效百分数累计

的。没有孩子的人的有效百分数是27.7%,只有一个孩子的是16.2%。累积百分数栏中的第一个数是以有0个孩子的有效百分数开始的。第二个数是27.7加上16.2(只有一个孩子),其值为43.9%,再加上有两个孩子的有效百分数(26.6),累积结果为70.5%(27.7+16.2+26.6=70.5)。你可以说,70.5%的回答了这个问题的人有至少两个孩子。注意,累积百分数栏中的最下面一个数字是100%,因为所有的有效百分数都已经累加上了。因此,它应该和有效百分数的总计相等。计算累积百分数时没有用百分数,因为累积百分数只基于回答了该问题的人,而不是基于接受此次调查的所有人。

使用累积百分数的前提条件是,数据必须要么是定序量,要么是定距/定比量。对没有顺序意义的定类数据,使用累积百分数没有任何意义。试想,对"你驾驶什么样的车"这个问题的可能回答(值),如本田、丰田、福特、雪弗莱等中,你能说有25%的人有一部丰田或少于一部吗?这种说法已经包含了顺序之义,但定类量数不具备顺序意义。

得出结论,解释结果:对该问题的回答看上去存在足够的变异,因此能保证随后进行更深入的数据分析。虽然大多数回答者有两个或两个以下孩子,但无孩子的人的百分比与有两个孩子的人的百分比几乎相当,所以,你有足够多的回答者,如对无子女的人和有子女的人作个比较。

比如,如果我们有20世纪20年代或20世纪60年代所调查的结果,我们便可以比一比,看今天的人要的孩子是不是比以前少了些。我们或许会发现,小家庭在当代美国社会更为普遍。

条形图

图1.4用条形图直观地展示了以上例子中的调查结果。

找出变量:虽然变量"孩子数量"是个定距/定比量,但它还是离散数据。如果这个变量是连续性的定距/定比量的话,直方图将是更好的选择,因为不同条形间的接触表明不同类别是

30
20
10
0
百分比/%
0 1 2 3 4 5 6 7 8或以上
孩子数量/个

图 1.4

连续的。由于不可能有诸如2.3或1.8个孩子这样的数据,条形图能更准确地描绘出该定距/定比变量的离散性。

解读图表:由于0和2是条形图上两个最高的条形,我们可以清楚地看到,没有孩子和有2个孩子的人的数量最多。尽管图上没有注明确切的人数,我们仍然可以通过这个图快速了解这个变量的数据分布情况。我们可以看到,大部分人的孩子并不多,最典型的是有一个孩子、两个孩子或者三个孩子。很多人都没有孩子,因为此次调查的对象也包括了没有结婚的人。

得出结论,解释结果:我们再一次得出这样的结论:当代美国家庭倾向于少要孩子。这个图表并未对此现象作出解释。这些是描述性数据,并不是用来解释或预测的。

描述性统计量

另一种描述变量及其反应结果的方法是用统计量。譬如说你对人们什么时候结婚,为什么结婚的年龄不同感兴趣,想了解更多这方面的信息,那么,在做进一步数据分析前,你必须确定调查表中的项目正是你样本中的变量。表1.6给出了一个频数表的统计量,这个频数表显示了《普通社会调查》第一次结婚时的年龄的调查样本。

表1.6 统计结果

第一次结婚时的年龄		
人数	有效人数	1 202
	缺失人数	298
均值		22.79
中位数		22.00
众数		21
标准差		5.033
全距		45
最小值		13
最大值		58
百分位数	25	19.00
	50	22.00
	75	25.00

来源：GSS 1993

找出变量：首先，在分析数据之前要知道变量是在哪个水平测量的，这一点很重要。有时候，研究者以定序量的方式提供年龄数据。如：(1)20 岁以下，(2)20～30 岁，(3)31～40 岁，(4)40 岁以上。这种年龄数据不适合计算均值和标准差。不过，在本调查中，年龄是定距/定比量，因此适合计算百分位数、均值和标准差。

解读图表：有 1 202 名被调查者报告了他们第一次结婚时的年龄，298 名没有报告。这 298 名中既包括没有结婚的，也包括已经结婚却拒绝回答或者由于忘记而不能回答这个问题的人。这些回答了问题的人的算术平均年龄(均值)为 22.79 岁，或约 23 岁。

根据中位数，一半的被调查者(601 名)在 22 岁前结婚，另外一半在 22 岁后结婚。最多的人第一次结婚时的年龄(众数)是 21 岁。若要知道 21 岁结婚者占样本的确切百分数，还需要参看频数表。根据这个问题的频数表(这里没有提供)，21 岁结婚者占样本人数的 11%。切记，不能说大部分人在 21 岁结婚，21 岁只是最常见的结婚年龄。毕竟，89% 的人不是在 21 岁

时结婚的。

如果均值、中位数、众数都相同的话,分布将会呈正态曲线型。不过,该例中的均值稍大于中位数和众数,这表明有些人第一次结婚时的年龄非常大。根据频数表提供的完整数据,至少有1人第一次结婚时的年龄是13岁,还有1人第一次结婚时的年龄竟然是58岁!全距45岁(58-13=45),而标准差只有5岁,这告诉我们,尽管应答者中有一些极端情况,大部人第一次结婚时的年龄可能都是在23±5岁这个范围内。

百分位数告诉了我们更多的分布信息。这里,每1/4处提供一个百分位数,这样的百分位数又叫四分位数。注意,50%处的百分位数与中位数相等,也应该相等。百分位数显示,样本中有25%的人在19岁或19岁以前结婚,75%的人在25岁或25岁以前结婚。或者反过来说,样本中有25%的人在25岁或25岁以后结婚。百分位数可以用来计算四分位数之间的跨距(例如:第75百分位和第50百分位之间的跨距为25-22=3岁)。

得出结论,解释结果:调查问卷中的这个项目看起来是一个在年龄上有一定分散度的变量,适合做进一步的数据分析。即使你可能不认识很多在年仅23岁时就结婚的人,但是要记住,这些数字是汇总数据。也就是说,它是1 202个人的总的情况。为了对第一次结婚年龄有更好的了解,可以根据性别(男性和女性谁结婚得更早)、国家区域、宗教背景、教育水平以及其他一些重要变量对数据进行分类,这样能提供比笼统的总体描述统计量更多的信息。但是,为了做这些分析,首先确定你是否有所需的变量,这一点很重要。例如,如果样本中90%的人来自中西部,那么被调查者来自哪里对于了解结婚年龄的地域差异并不是一个有用的变量。

练习框 1.4

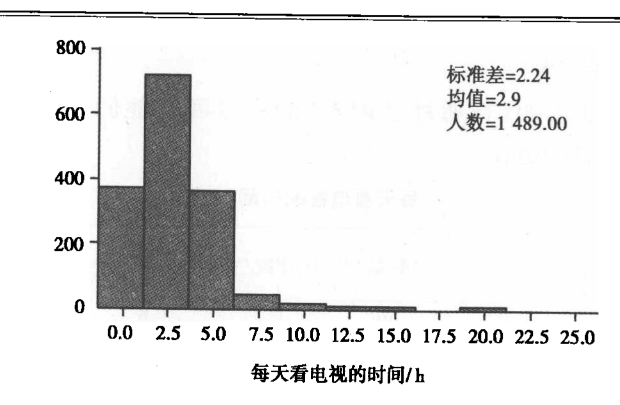
要求参照《普通社会调查》的样本回答他们一天大概看几小时的电视。

每天看电视的时间

		频数	百分数/%	有效百分数/%	累积百分数/%
有效	0	56	3.7	3.8	3.8
	1	311	20.7	20.9	24.6
	2	422	28.1	28.3	53.0
	3	282	18.8	18.9	71.9
	4	183	12.2	12.3	84.2
	5	95	6.3	6.4	90.6
	6	67	4.5	4.5	95.1
	7	10	0.7	0.7	95.8
	8	32	2.1	2.1	97.9
	9	2	0.1	0.1	98.1
	10	15	1.0	1.0	99.1
	12	8	0.5	0.5	99.6
	16	2	0.1	0.1	99.7
	20	2	0.1	0.1	99.9
	22	1	0.1	0.1	99.9
	24	1	0.1	0.1	100.0
	总计	1 489	99.3	100.0	
缺失	NA	11	0.7		
总计		1 500	100.0		

统计结果

		每天看电视的时间
人数	有效	1 489
	缺失	11
均值		2.90
中位数		2.00
众数		2
标准差		2.238
全距		24
百分位数	25	2.00
	50	2.00
	75	4.00



问　题

1. 变量是什么,它是在哪个水平上测量的?
2. 哪些统计量相关、有用?
3. 百分数、有效百分数和累积百分数告诉了我们什么? 有多少人回答了这个问题? 样本中共有多少人?
4. 用文字叙述频数表的内容。例如:有效百分数列中的6.4告诉了我们什么? 你可以这样开始:6.4%的……
5. 解释均值、中位数、众数、全距、标准差和百分位数。
6. 为什么这个例子适合用直方图? 解释该直方图。
7. 用日常生活语言表述,这个调查告诉了我们美国人看电视方面的什么信息?
8. 对于这个结果的可能解释是什么? 有没有什么理论或先前的研究可以引导你解释这些发现?

描述性统计量、统计表和统计图部分小结

作为数据分析的第一步,我们要确定调查问卷中有多少个项目实际是变量,对于进行更复杂的数据分析有用,这一步很重要。此外,我们还要通过利用各种直观的统计图表、频数表以及各种趋中统计量和离中统计量,以描述被调查者的特征,

凸现某些关键性的发现，讨论所呈现出的一些态势。

当我们在阅读公众意见调查结果，解读大众传媒中常用的饼图或条形图时，一是要了解问题是怎样表述的，二是要了解变量是怎么测量的，这两点必不可少。诚如 Darrel Huff（1954）在其经典著作《怎样用数据说谎》中证明的那样，不同的用词可以导致许多截然不同的结果，不正确地使用统计量或统计图表很容易造成错误的理解。

尽管呈现一元变量的信息很常见，但在学术研究、新闻报道和调查中更典型的做法是呈现二元数据信息。我们很少只想知道每个变量自身的信息，更多时候，我们想知道两个变量、三个变量甚至更多个变量之间的关系，从而帮助我们对结果进行解释。对于了解复杂行为和意见，二元分析或多元分析至关重要。下一章，我们将讨论二元数据及其常用图表的解读问题。

第2章　理解表格

研究的中心目的之一就是描述、说明或者预测两个或多个变量之间的关系。对研究中每个变量的分布情况像第1章讨论的那样进行描述,这固然是很好的做法,但是,分析两个或多个变量之间的关系更是必不可少。例如,找出一个政党偏爱性的分布状况,发现一次投票选举行为中男性和女性的数量分布很重要,但关系更为密切的是分析哪个性别的投票者更可能支持或不支持某个政党候选人。

揭开这些二元关系的方法之一,就是同时比较两个变量,然后编制出一个表格,把两个变量的分布情况都表示出来。这种表格叫做交叉表或列联表,是大众出版物、学术期刊和专业报告中最常用的展示数据的方法之一。本章将解释如何解读表格,理解数据,然后再把表格中的信息用日常语言清楚地表述出来。

定 义

■ **交叉表或列联表**:这是一种同时展示两个变量的频数分布信息的统计表。我们假设,一个变量(因变量或结果变量)取值的分布伴随或者取决于另外一个变量(自变量、预测变量或原因变量)取值的分布。

■ **行、列和单元格**:这三个概念是理解交叉表的关键。交叉表的形状和其他开摊式表格一样,就是像 Excel 电子表格和 SPSS 统计软件的数据界面那样。交叉表就像剧院里的座位一样。试想象一下,在剧院里找座位的情景:你先沿着走道走到你座位所在的排(行),然后沿着这一排横穿过去,走到你座位所在的列,这个排列交叉处就是你的座位。如果你向前或向后看就可以发现,前后的人都和你坐在同一列上。换句话说,你的行号或字母表示左右横穿的地点,你的座位号就是座位所在处的行号(字母)和列号。如果你的座位号是 M24,那么,你的座位就在 M 行的第 24 列,这表示你的座位在 N24 的前面,在 L24 的后面。行从左到右(水平方向),列则像柱子一样,从上到下(竖直方向)。交叉表通常用构成表格的行数和列数来描述,所以,一个有 3 行 4 列的表就是一个 3 ×4 表格。表格中行列交叉形成的小格子叫做单元格。你可以通过单元格所在的行数和列数来说明单元格所在的位置,例如:第三行第二列。

假 定

■ 交叉表最适合于定类变量或离散性定序变量。行数或列数太多不便于阅读,这就使得交叉表不是描述连续性定距/定比变量的最好方法,因为这类变量的类别或取值太多。

■ 为了较容易解释,很多研究者在制作交叉表时都用自变

量作列,因变量作行。不过,出于对表格大小的考虑,有时也会用自变量作行,因变量作列。

■ 频数与百分数一起呈现时,表格最容易读。为了比较自变量的不同类别对因变量结果的影响,自变量的类别或取值加起来必须等于100%,尤其当变量的每个类别中的应答数量不等时,这一点尤为重要。当列中放的是自变量的类别或取值时,每列的类别加起来必须等于100%。如果自变量放在行中,每行中的类别加起来也一样要等于100%。

取自大众出版物的真实数据

让我们从解读典型的交叉表开始。当需要展示民意调查的结果时,日报上常会出现如表2.1所示的表格。

表2.1 不同状况下对于修改宪法的观点

	先问对同性婚姻的看法/%	后问对同性婚姻的看法/%
支持修改	48	54
反对修改	47	41
不知道	5	5
	100	100

来源:《洛杉矶时报》(2004年6月)"对于修改美国宪法关于婚姻只限于男女结合,并防止各州承认同性婚姻的合法性,您是支持还是反对?"

Source: Los Angeles *Times*(June 2004) "Do you favor or oppose an amendment to the U. S. Constitution that legally defines marriage as a union between a man and a woman only, and would prevent states from legally recognizing same-sex marriages?" Question rotated with baseline measure of support or opposition to gay marriage.

简单二元统计表

表2.1展示了《洛杉矶时报》2004年6月的一次民意调查的结果。这次调查是关于支持还是反对"修改美国宪法关于婚姻只限于男女结合,并防止承认同性婚姻的合法性"的问题。

对于一些被调查者，在提这个问题之前，先问他们对于同性婚姻的看法；对于另外一些被调查者，先让他们回答这个问题，然后再询问他们对于同性婚姻的看法。所以，这组数据不仅表明了人们对于修改宪法的意见，还说明了问题提出顺序的重要性。

找出变量：第一步是先找出交叉表里都有什么变量，并弄清为什么交叉表适合这些变量。通常，把每个变量的类别标志放在表格外面，或者放在表格的第一行和第一列。在这个例子中，顶行是“提问顺序”这个变量的两个类别（或取值）：“先问对同性婚姻的看法”和“后问对同性婚姻的看法”。这是一个只分为两类（有两个取值）的定类变量。

这个表的行包含了是否支持修改宪法这个主要变量的三个取值（支持、反对、不知道）。这也是一个定类水平的量。因此，两个变量都适合用交叉表和百分数。

按照常理，自变量各类别加起来应该等于 100%。所以这个表格告诉我们，先问对同性婚姻的看法还是后问对同性婚姻的看法是自变量。

解读图表：注意，这里每列加起来都等于 100，所以，这些数字是百分数，而不是原始的频数，除非刚好有 100 个人，在征求他们关于修改宪法的意见之前，先询问了他们对同性婚姻的看法，而另外 100 个人则刚好相反，先征求他们关于修改宪法的意见，然后再询问他们对同性婚姻的看法。只有在这种情况下，原始数据才会与百分数相同。不过，本交叉表中并没有提供原始数据（频数）。根据每个自变量的值加起来必须是 100% 这条规定，这个交叉表还告诉我们，提问顺序是自变量。应答者是否支持修改宪法取决于他们听到问题的顺序，所以关于修改宪法的意见是因变量或结果变量。表中的标志也告诉我们，关于修改宪法的意见“因语境而异”，这使得应答者的意见成为因变量，态度取决于语境。

换句话说，人们的反应取决于问题的顺序。另一种说法是，问题提出的顺序导致了人们关于修改宪法的态度。由于这

是从一大群人中得到的结果,我们不可以推广到样本中的任何一个个人,或者推广到全体美国人,除非能够确保这个样本是从总体中随机抽取的。

这个交叉表告诉我们,被先询问他们对于同性婚姻看法的人中,有48%的人支持修改宪法,被后询问他们对于同性婚姻看法的人中,有54%的人支持修改宪法。与先听到同性婚姻问题的组相比,先听到修改宪法问题的组中看起来有更多的人支持修改宪法。这个交叉表并没有说,支持修改宪法的人当中有48%的人先听到关于同性婚姻问题,54%的人后听到关于同性婚姻问题。这一点很关键,很多人都会弄混。再举一个例子,不能仅仅因为大学中主修社会学的学生中有65%是女生,就说女生中的65%是主修社会学的。解读列联表时,应该慎重考虑这两种不同的意思。

要记住,百分数是根据每个自变量不同类别中的人数计算出来的。解读统计表时,很重要的一点是要在自变量的不同类别间进行比较。这就是为什么自变量的每个取值或类别加起来必须是100%,变量不同,取值之间方可进行比较。

得出结论,解释结果:一个说得过去的结论是,调查问题的提问顺序会导致民意调查结果的不同。先问被调查者对同性婚姻的看法,结果,他们中支持和反对修改宪法的人数大致相同,这与那些后被问对同性婚姻看法的组形成了对照。后被问的组更倾向于支持修改宪法。这些数据间的差异是否在统计学上显著,也就是说,得到这个数据的概率是否小于5%($p < 0.05$),在大众传媒出版物上通常并不给出。但是,学术研究要求给出统计分析的显著性水平。此外,也不能像我们这里做的这么简单,仅凭看统计图表就作出结论。第3章将讲述如何解读交叉表中出现的典型统计结果。

在作最后结论之前,很重要的一点是要弄清楚该民意调查样本的抽样误差。《洛杉矶时报》典型的民意调查抽样误差是±4%。这告诉我们,由于调查的样本大小的局限,把样本调查的结果推广到全体国民时,必须考虑误差。也就是说,如果所

有被调查者中选“是”的有 70%，那么全体美国国民中选“是”的真实百分数可能在 66% 到 74% 之间（70 - 4 = 66，70 + 4 = 74），并不正好是 70%。这里展示的交叉表显示，两组间有 6% 的差异（48% 和 54%），所以我们很可能得出这样的结论（尽管我们还不能在统计学上得出这样的结论）：问题的顺序对人们是否支持修改宪法可能会有影响。

为什么会有这样的差异呢？我们还不能从这个交叉表中确定答案。也许，那些先听到修改宪法问题的人，也根据这个问题表达了他们对同性婚姻的反对，而那些有机会先表达他们看法的人现在能够全心关注修改宪法是否合适。研究一下有关问卷设计的文献，也许我们能找到一些关于问题顺序影响调查结果的证据。

多变量交叉表

另外一种典型的呈现数据的统计表如表 2.2 所示。该表显示的是一项皮优大众与传媒研究中心（Pew Research Center for the People and the Press）关于广播电视节目在不同政党成员之间受欢迎程度的调查结果，调查样本是 3 000 个美国成年人，调查时间是 2004 年 6 月。这个交叉表与前面那张交叉表稍微有点不同，即它的每行加起来并不是 100%。该表每个单元格报告的仅仅是“常规”的反应结果。每列代表不同的变量，这些并不是一个变量的 13 个类别或取值，因为它们并不是互不相容的类别。人们可以选不止一个他们获得新闻的途径。这个交叉表中的每一行都是一个更大数据表的总结，这些数字代表了 13 个不同交叉表的结果。

表 2.2 党派成员的不同新闻来源

经常收看频道或节目	占总数的百分比/%	共和党人所占比重/%	民主党人所占比重/%	无党派人士所占比重/%
福克斯新闻频道	25	35	21	22
CNN	22	19	28	22
MSNBC	11	10	12	12
CNBC	10	9	12	9
NBC 每晚新闻	17	15	18	19
CBS 晚间新闻	16	13	19	17
ABC 世界新闻	16	15	20	12
NPR	16	13	19	17
新闻小时	5	4	5	5
O' Reilly Factor	8	16	3	6
Rush Limbaugh	6	14	2	4
Larry King	5	5	6	3
每日播报	3	3	3	3

来源：Pew Research Center for the People and the Press（2004 年 6 月）

找出变量：这张统计表中的自变量是政治派别，这是个定类变量，表中有共和党人、民主党人、无党派人士三个类别（取值）。因变量是 13 个广播电视新闻节目和新闻频道，有 CNN 和“每日播报”等。调查的问题是，这些政党成员是否定时、有时、很少或从不收看或收听这些节目。因变量是定序变量，表中仅仅显示了选“定时”的那些人的百分数。假设是：收看或收听不同的新闻节目取决于政治派别。

该交叉表展示出的基本是 3 个自变量与 3 个因变量之间不同双向交叉表的结果。在那 13 个双向交叉表中，政治派别都是自变量，收看或收听某一新闻节目的频数/率都是因变量。表 2.2 中每行列出的仅仅是“定时”类的结果。用第一列举例来说明，整个样本中 25% 的人说他们定时收看福克斯新闻频道，75% 的人说有时、很少或从不收看这个频道。由于其他的数据没有给出，因此我们不知道“有时”或“没有”看福克斯新闻频道的人的确切数据。原始统计表看起来可能有点像表

2.3,不过给出的是“有时”、“很少”、“从不”这三个类别的实际百分数。

表 2.3

福克斯新闻频道	总数	共和党人	民主党人	无党派人士
经常	25%	35%	21%	22%
有时	××	××	××	××
很少	××	××	××	××
从来没有	××	××	××	××
总计	100%	100%	100%	100%

解读图表:皮优大众与传媒研究中心的报告明确显示:共和党人更倾向于定时收看福克斯新闻频道,还有 Bill O' Reilly 和 Rush Limbaugh;相比之下,民主党人则更欢迎 CNN、CBS 和 NPR。

我们可以说,共和党中有 35% 的党员定时收看福克斯,而民主党中只有 21%。注意,这并不是说收看福克斯节目的观众中 35% 是共和党人。从表 2.3 中是得不出这些数据的。

“总计”行告诉我们,在这 3 000 人的调查样本中,总的情况是,大家更喜欢从福克斯新闻、CNN、NBC、CBS、ABC 和 NPR 获取日常新闻,与对政党的忠诚度无关。在许多统计方案和出版物的统计表格里,总数经常出现在最后一列,而不是第一列。解读表格的方法之一是,先看总数,然后再看每组的数字是大于、小于还是与总数相同。例如,总人数中的 5% 定时收看 Larry King 节目,这与共和党人(5%)、民主党人(6%)和无党派人士(3%)大致相同。换句话说,党派不能帮助我们预测谁会看 CNN 上的 Larry King 节目。然而,22% 的调查对象收看 CNN,与 28% 的民主党人和 19% 的共和党人收看 CNN 形成对比。显然,民主党比全部样本的 22% 要高,而共和党人收看 CNN 的比例则低于这个比例。总数对作这类解释很有帮助。

得出结论,解释结果:这里似乎有个模式,对于传统的新闻台——福克斯——和传统的主持人,即 Bill O' Reilly 和 Rush

Limbaugh,共和党人比民主党人和无党派人士更常定期收听或收看。民主党人更喜欢 CNN 和 NPR。也就是说,似乎政党党派与选择新闻节目之间有关系。

党派为什么会造成对新闻节目选择的不同呢?表中并没有说明。对不同节目的语言、偏向和形象做探索性的内容分析,也许会找出一些理由来。这些发现例证了一组数据可能产生出新的研究问题和方法,有待我们作进一步的研究。

练习框 2.1

2003 年,皮优大众与传媒研究中心在美国全国范围内对 1 200 个成人做了一个调查。调查中问了这样一个问题:"为了消除过去的歧视,你对旨在帮助黑人、妇女和其他少数民族得到更好的工作和教育的希望行动计划是持反对意见,还是持支持意见?"另一个问题是:"总而言之,你觉得这个旨在增加大学中黑人和少数民族学生数量的希望行动计划是一件好事还是坏事?"下面是两个交叉表,报告了白人男性和白人女性的回答结果。

对希望行动态度的性别差异

计划内容		
帮助黑人、女性和其他少数民族得到工作和教育	白人女性/%	白人男性/%
支持	66	48
反对	26	41
不知道	8	11
	100	100
增加大学生中的少数民族		
好事	60	48
坏事	28	43
不知道	12	9
	100	100

问　题

1. 自变量和因变量是什么？它们的测量结果是什么水平上的（定类、定序、定距/定比）？
2. 要检验的假设是什么？
3. 为什么这里适合用交叉表？
4. 百分数说明了什么？用文字描述表格内容，如：第二列中的 26 说明了什么？你可以这样开始：26%……
5. 用日常生活语言叙述这次皮优调查的结论。
6. 你是否有足够的理由得出这里存在或者不存在“性别差异”的结论？还有什么别的东西能帮助你得出结论吗？
7. 对于该结果可以说得过去的解释是什么？有没有什么理论或先前的研究指引你解释这些发现？

取自学术文章的真实数据

在学术文章中，经常用统计表格表示两个或更多变量之间的关系，并带有关于所描述的关系是否达到显著性水平的有关统计量。第 3 章将介绍这类统计量，因此，我们还是首先展示表格在文章中是如何应用的，并学着解释这些表格。

简单的二元统计表

Stein（1998）报告了一项关于得克萨斯州人在州长选举方面的习惯的研究。他对那些在选举日当天投票的人和事先投票的人之间的人口统计学特征差异很感兴趣。在得克萨斯州，人们可以从选举日前 3 个星期起到指定的地点投票，将近 30% 的人提前投票。表 2.4 展示了部分结果。

表 2.4 提前投票和选举日投票选民的部分特点(表中数据为%)

变 量	选举日(N=1 362)	提前(N=1 531)
党派		
坚定的共和党人	24.8	26.9
共和党人	6.9	4.6
不坚定的共和党人	2.5	1.7
无党派人士	26.5	27.0
不坚定的民主党人	2.4	1.5
民主党人	8.0	6.2
坚定的民主党人	29.0	32.0
性别		
男性	47.6	52.1
女性	52.4	47.9
教育程度		
高中以下	7.1	7.5
高中毕业	20.0	22.3
受过一定程度的大学教育	33.1	30.0
大学毕业	24.3	22.2
研究生学历	15.5	18.0

来源:Stein (1998)

找出变量:这个统计表可以称作“列百分数”表。我们知道,这类表格要经过精心设计,使变量各类别(取值)加起来为100%,以便进行比较。这类表格告诉我们(不用我们自己亲自加),比较是在提前投票的人和在选举日投票的人之间展开的。我们可以说变量是“投票时间”,它有两个类别或两个取值:“选举日”或“提前”。

构成行的是三个不同的人口统计学变量,分别是政党派别、性别、教育程度。所以,这张统计表实际上是由三个分开的列联表合并而成的。

与该研究有关的假设可能是:什么特征预示了谁会提前投票?教育程度会影响投票时间吗?提前投票取决于党派、教育程度或性别吗?投票时间会不会影响选举结果呢?换句话说,根据研究重点的不同,投票时间既可以是因变量,也可以是自

变量。尽管这些可能是本研究要探究的部分问题,但是,第一个目标还是对提前投票者和当天投票者的特征进行了描述。

不过这向我们提了个醒:在设计表格之前,我们首先必须清楚研究的目标和研究的问题。在这个例子中,研究目标是对投票时间两个类别上的情况进行比较。所以,投票时间的两个类别是列,而且各自分项上的值加起来等于100%,这样就便于进行比较。有些情况下,确定因变量和自变量并不是那么重要,本例就是其中之一。因为我们还不能说提前投票发生在教育程度、性别或党派之前,也不能说提前投票造成了或者预测了教育程度、性别或党派。这里的目标仅仅是在两个投票时段间进行比较。这个目标决定了百分数的方向,在这个例子中,列表示不同投票时段投票情况的对比。

在这个表格中,变量要么是二分的(即两个类别,例如选举日或提前、男性或女性),要么是定序离散的(党派和教育程度都有一定的顺序,所以它们不再是定类变量)。可见,表中的所有变量都适合于交叉表。

解读图表:这项研究的零假设可能是:提前投票的人和在选举日投票的人之间没有什么差异。让我们先来看一下关于党派的第一个交叉表。注意,两个样本的人数非常接近,选举日投票的人数是 1 362,提前投票的人数是 1 531。由于两个样本的人数并不正好相等,为了对两个类别的情况进行比较,表中用了百分数。这张统计表告诉我们,在选举日投票的人当中,有 24.8% 的人是坚定的共和党人,提前投票的人当中有 26.9% 的人是坚定的共和党人。尽管有些差异,但看起来并不是很大。事实上,如果你把所有的共和党人加起来(从坚定的到不坚定的)你会看到,在选举日投票的人中有 34.2% 是共和党人,提前投票的人中有 33.2% 是共和党人。这些数字在统计学上似乎并没有什么差异。

相似地,选举日当天投票的人中有 39.4% 是民主党人,提前投票的民主党人是 39.7%,两者几乎一样。换句话说,不同时间投票选民的政治面貌分布情况基本上一样。回想一下,这

些数据并没告诉我们投票日投票和提前投票的民主党人占全部民主党人的百分比。选举日当天,所有注册投票的人中有39.4%是民主党人是一个意思,所有注册的民主党人中有39.4%在选举日投了票是另外一个意思。不过,从这个表格中我们无法知道民主党人中到底有百分之几在选举日投了票。这些数字所告诉我们的是:在选举日投票的人中民主党人所占的比例与提前投票的人中民主党人所占的比例。如果你想要比较民主党人和共和党人的选举习惯,那你需要计算不同的百分数。这样的话,党派可以作列,而选举时间的不同类别可以作行。这样,你就可以说百分之几的民主党人提前投票,百分之几的共和党人提前投票,对两党的情况进行比较很方便。

看起来提前投票的人和在选举日投票的人在性别上确实有很大的差异。提前投票的人中有47.9%是女性,在选举日投票的人中却有52.4%是女性。这些差异可能在统计学上具有显著性。切记,这些数据并没有告诉我们,提前投票的女性和在选举日投票的女性各占全部投票女性的百分之多少。

在不同教育程度上对提前投票者和选举日投票者作比较,结果看起来并没有多少差异。在选举日投票的人中有15.5%的人完成了一定程度的研究生学习,提前投票的人中则有18%的人具有研究生学历。同前面一样,这个表格并不能说明具有研究生学历的人当中有18%提前投票。这里,我们不能对根据选举时间对教育水平的五个等级进行比较。要想对教育的分类进行比较(把它作为自变量),每一类就必须重新计算,总计要为100%。

得出结论,解释结果:从这些数据我们可以得出结论:男性提前投票的比例要比在选举日投票的比例更大些。除此之外,在党派和教育程度上进行比较,不同选举时间的投票比例并没有什么不同。提前投票的民主党人和共和党人的百分数与在选举日投票的百分数一样。发表这些数据的文章中,还对其他变量做了详细的分析,而且还做了一些更为复杂的统计分析。文章在分析实际选举情况时用了党派变量,也用了这些人口统

计学变量。

为什么在性别上有差异,而在党派和教育程度上没有差异呢? 关于这个问题有几种可能的解释。这里呈现出的数据并不能帮助我们解释这些结果,因此,建立在以前研究结果和理论之上的猜测很可能就是我们理解这些发现的引导。对政治和政党感兴趣的人更倾向于提前投票,不那么感兴趣的人则倾向于当天投票。是不是得克萨斯州的男性要比女性更多地参与政治活动并且有更大的政治倾向性? 另一个可能解释与工作有关。由于工作上的方便使得某个性别的人提前投了票或者由于工作原因而不去投票。其他的研究或者进一步的研究可能会解释男女投票模式的差异。

应该记住的一点是:一旦你得到一组结果,你的任务才刚刚开始。像这组数据那样的描述性发现,往往会引出另外一些更为复杂的研究问题。

多元交叉表

比较两个变量往往只不过是一个好的开始,通常还需要加入另外一些变量来作分析。表 2.5 就朋友冲突的根源与持久性的看法,对 73 个男生和女生(四、五年级)作了比较。问学生的问题是:“究竟为什么会与朋友发生争吵?”他们的回答分成两个变量:冲突的根源和冲突的持久性和暂时性。

找出变量:表里涉及三个变量:第一个是应答者的性别(男生或女生),是定类二分变量;第二个是冲突的持久性或暂时性,同样也是定类二分变量;第三个是归因,是定类变量。归因涉及人或关系的特性(归因是由于友谊或人的特性不同,譬如“人是不同的”)、交往条件(朋友间产生的目标冲突)、个人特性(他人或自己的行动或性情,譬如“她心情不好”)以及外部特性(关系之外的其他东西)。在这个例子中,归因是因变量,性别是自变量,冲突的持久性/暂时性是第三个变量,通常叫做控制变量。

定类变量最适合于交叉表和百分数分布。

表 2.5 朋友间冲突的持久性和归因

	持久(25 人)		暂时(48 人)		总计(73 人)
	女生 (16 人)	男生 (9 人)	女生 (23 人)	男生 (25 人)	
人或关系的特性	16(100.0)	8(88.9)	6(26.1)	3 (12.0)	33 (45.2)
交往条件	0(0.0)	1(11.1)	11(47.8)	14 (56.0)	26 (35.6)
个人特性	0(0.0)	0(0.0)	6(26.1)	7(28.0)	13 (17.8)
外在特性	0(0.0)	0(0.0)	0(0.0)	1(4.0)	1(1.4)

注:括号里的数字表示列的百分数。

来源:Joshi and Ferris ,2002

解读图表:一个有三个变量的表格,可以沿不同的方向来读。我们还是先看最后一列的总计,这些是本调查发现的整体总结。几乎一半的孩子(45.2%)把与朋友的冲突原因归咎于人或关系特性,也就是归咎于人与友谊的本质。超过三分之一(35.6%)的孩子认为,形成冲突的原因是朋友相处过程中的交往条件。

现在再把这些整体数字与控制变量不同类别中关于男生和女生的发现进行比较。我们可以看到,认为和朋友的冲突是持久的孩子,在性别上几乎没有差异。男生和女生都把持久性的冲突归因于友谊和人的特性(100%对88.9%),这比整体结果(45.2%)高出许多。

但是,那些认为冲突是暂时的孩子中,男生似乎比女生更多地把他们和朋友发生冲突的原因归咎于交往条件(56%对47.8%)。此外,和总体结果相比,男生和女生都更倾向于把暂时的冲突归因于个人特性(28%和26.1%对17.8%)。

得出结论,解释结果:男生稍微更倾向于把冲突的形成,尤其是暂时性冲突的形成归因于交往过程中的条件,而女生则更倾向于用友谊和人的本质来解释冲突的形成。然而,无论是男生还是女生,对于冲突是持续的还是暂时的都有不同的解释:

持久性冲突是由友谊或人的本性造成的，而暂时性冲突则归因于具体的交往过程或具体参与交往的人。

有些发展理论认为，随着孩子的长大，他们会更多地在交往情景中寻找解释，更少地在人的天性中寻找解释。也许，当孩子把与朋友的冲突更多地看做是暂时性的时候，他们便开始注意到有多少冲突取决于交往情景。像往常一样，关于这些看法的确立，还需要进一步的研究。

练习框 2.2

Looker 和 Thiessen (1999)采访了 1 700 名 17 岁的加拿大孩子，问他们心目中对于女性工作和男性工作的形象问题。下表是问及他们是否愿意考虑从事通常属于另一性别的工作时的回答总结。

不同性别青年对于从事通常属于另一性别的工作的态度

您会考虑从事一项现在大部分从业人员是男性/女性的工作吗?					
	一定会	也许会	也许不会	一定不会	数目
应答者性别					
男性	9%	42%	42%	7%	555
女性	32%	48%	18%	3%	633
总计	21%	45%	29%	5%	1 188

来源:Looker and Thiessen (1999)

问　题

1. 哪个是因变量? 哪个是自变量? 它们的测量水平分别是什么?
2. 拟检验的假设是什么?
3. 行或列加起来等于 100% 吗? 为什么这一点有助于解读数据?
4. 用文字描述这个表格告诉了我们些什么。例如:第一行、第三列中的 42 是什么意思? 请用文字表达。你可以这样开始:42% 的……

5. 用日常生活语言叙述本研究的结论。
6. 你能否得出结论说,有或者没有性别差异?还有什么可以帮助你作出结论?
7. 对于这里的结果,可以说得过去的解释是什么?有没有什么理论或先前的研究指引你解释这些发现?

SPSS 结果输出

为报纸或学术期刊写稿子时,作者绝不是把统计软件输出的结果简单地搬到文章中就一了百了。你在文章中看到的也许和统计软件输出的直接结果不一样。所以,这里选了一些SPSS 软件的结果输出,请你解释。

二元表格

表2.6 展示的是些交叉表数据,用《普通社会调查》(GSS)的结果来检验种族与对RAP 音乐态度之间的关系。这是从全美国18 岁以上的成人中随机抽取的样本。

表2.6 RAP 音乐*应答者种族交叉表

			应答者种族			
			白人	黑人	其他	总计
RAP 音乐	喜欢	人数	117	52	17	186
		占所属种族的百分比	9.7%	32.3%	24.6%	13.0%
	复杂情感	人数	205	41	20	266
		占所属种族的百分比	17.1%	25.5%	29.0%	18.6%
	不喜欢	人数	879	68	32	979
		占所属种族的百分比	73.2%	42.2%	46.4%	68.4%
总计		人数	1 201	161	69	1 431
		占所属种族的百分比	100.0%	100.0%	100.0%	100.0%

找出变量:在SPSS 统计软件输出的交叉表中,很清楚地标

出了每个变量及其取值。列是“种族”,分成了三个类别或三个取值:白人、黑人和其他。由于完成调查的拉丁人和亚洲人太少,所以不能把他们各自看做单独的一类,而只能和其他少数民族合并成“其他”。行是对 RAP 音乐的态度,包括“喜欢”、“复杂感情”和“不喜欢”。

总计列就是所有应答者的汇总,也就是每个应答者(不管是什么种族)对 RAP 音乐各种态度回答结果的总和。没有把回答结果再根据种族分组,总数就是 1 431 个应答者对调查问卷中这个项目(13% 喜欢等)回答结果的一元描述统计结果。有些表格还显示每列的总计,放在最后一行(本例中 1 201 个应答者是白人)。包含总计的行或列叫做边际,因为它们不在表格的中央,而在表格的边上。计算表格的大小时,这些边际行或列并不计算在内。所以在这个例子中,表格的大小为 3 × 3,即有 3 行 3 列的数据,不能包括任何包含“总计”的行或列。

由于列加起来是 100%,我们可以通过种族的三个类别或取值(自变量)来比较应答者对于 RAP 音乐的态度(因变量)。换句话说,我们要问的是对于 RAP 音乐的品味是否取决于种族。黑人、白人或其他是不是更喜欢 RAP 音乐?知道某人的种族能否预测其对 RAP 音乐的态度?

因为目的是根据不同种族来描述态度上的差异,又因为种族是定类变量,对于 RAP 音乐的态度是定序变量,所以,交叉表或列联表是展示结果的一个很好选择。

解读图表:通过最右边列中的总计,我们首先可以看到,大部分人(68.4%)不喜欢 RAP 音乐,剩下的人则对 RAP 音乐持有复杂情感或正面的情感。然而,当我们把种族分成三个类别时,不同的模式就出现了。白人中的大部分(73.2%)不喜欢 RAP 音乐,只有 9.7% 的人喜欢。这与少数黑人(42.2%)和其他人种(46.4%)不喜欢 RAP 音乐的比例形成了对照。

再看一下因变量的类别,每类的原始数据都转化成了百分数使其标准化。这也是为什么需要百分数的一个很好的例子。如果你仅仅有频数的话,你可能得出白人(117)比黑人(52)和

其他人种(17)更喜欢 RAP 音乐的结论。但是,在这个调查中,完成问卷的白人要多得多,所以,只有把原始数据转化成百分数之后,才能在人数不一样的组之间进行比较。117 除以 1 201 得出样本中 9.7% 的白人喜欢 RAP 音乐的结果。

还要注意,这些数据表明,9.7% 的白人喜欢 RAP 音乐,但并不是喜欢 RAP 音乐的人中 9.7% 是白人。如果你想找出喜欢 RAP 音乐的人的特性,那么,你就必须计算出行的百分数。例如,用 117 个喜欢 RAP 音乐的白人除以调查中所有喜欢 RAP 音乐的人(186),就得到调查中喜欢 RAP 音乐的应答者中有 63% 是白人。通过把这个数字与白人在整个人口分布中的比例作个比较你可以得出,喜欢 RAP 音乐的白人是否和白人在总人口中的多少成正比或反比。

得出结论,解释结果:把调查结果翻译成日常生活语言我们就可以说:尽管大部分人不喜欢 RAP 音乐,但是表明自己喜欢 RAP 音乐的黑人比白人多,实际上大概是 3 倍的关系(32.3% 对 9.7%);表明自己喜欢 RAP 音乐的其他人种也比白人多,只不过程度要比黑人低一些。

通过这个表格我们并不知道为什么在对 RAP 音乐的态度上会有种族差异。要解释这些结果,需要加入其他变量做进一步的分析。也许年龄是个因素,如果只分析 18 ~ 30 岁的人的话,所有种族类别中的大部分人都可能喜欢 RAP 音乐;也许即使在这个年龄段上,大部分人还是不喜欢 RAP 音乐,而且出现有规律的种族差异。关于音乐品味和种族差异的以往的研究能够帮助我们发展出理论,指引我们对研究发现作出解释。

多元统计表

一种更加复杂的分析涉及多个自变量和一个因变量。例如,试想我们对过去一年艺术博物馆的参观情况感兴趣。年龄可能是解释参观情况的变量之一,同样,性别也可能是。表 2.7 对来自《普通社会调查》的数据,就这两个变量同时进行分析。

表 2.7　不同年龄/性别去年参观博物馆/艺术馆双向表

应答者性别				年龄组				总计
				18～29	30～39	40～49	50＋	
男性	去年是否参观过博物馆或艺术馆	是	人数	46	66	65	68	245
			占该年龄组百分数	38.0%	41.8%	46.1%	31.2%	38.4%
		否	人数	75	92	76	150	393
			占该年龄组百分数	62.0%	58.2%	53.9%	68.8%	61.6%
	总计		人数	121	158	141	218	638
			占该年龄组百分数	100.0%	100.0%	100.0%	100.0%	100.0%
女性	去年是否参观过博物馆或艺术馆	是	人数	73	88	81	115	357
			占该年龄组百分数	46.5%	45.8%	49.1%	34.2%	42.0%
		否	人数	84	104	84	221	493
			占该年龄组百分数	53.5%	54.2%	50.9%	65.8%	58.0%
	总计		人数	157	192	165	336	850
			占该年龄组百分数	100.0%	100.0%	100.0%	100.0%	100.0%

找出变量：参观艺术博物馆是一个定类变量，而且是简单的二分变量（是/否）。是否参观取决于年龄（定序变量）和性别（定类二分变量）。拟检验的假设是：年龄和性别可以解释人们参观博物馆的情况。

解读图表：一个较大的交叉表可以分解成两个小交叉表来看，每个小表都是关于第三个变量的一个类别的。在这个例子中，第三个变量是性别。先看关于男性的那个表格，看看年龄与参观博物馆的关系。例如：38% 的 18 ~ 29 岁的男性去年参观过博物馆或艺术馆。对参观情况在年龄组间进行了比较，看男性的参观情况是否存在年龄差异。比起其他组的男性来，40 ~ 49 岁的男性看起来更经常参观博物馆（46.1%），50 岁以上的男性参观得最少（31.2%）。在每一个年龄组的男性中，大部分去年都没有参观过博物馆。

现在再来看女性的情况。相似地，最常参观的也是 40 ~ 49 岁年龄组，参观得最少的也是 50 岁以上年龄组。

那么不用年龄而用性别作解释又如何呢？这项任务可以通过比较同一个年龄组的男性和女性的参观情况来完成，如比较 18 ~ 29 岁组的男性和 18 ~ 29 岁组的女性。从交叉表中我们可以看到，18 ~ 29 岁的女性中，有 46.5% 的人去年参观过博物馆，而这个年龄组的男性中，只有 38% 的人去年参观过博物馆。事实上，在每个年龄组，女性参观博物馆的比例看起来都要比男性稍微高一些。这种差异在较年轻组内尤为明显，其他年龄组内的差异看起来并不大，或者说：在其他年龄组内，男性和女性参观博物馆的比例差不多。

得出结论，解释结果：我们是否就能够得出，在参观博物馆上存在年龄和性别差异？看上去年纪大点的人参观得最少，所以年龄可能是一个有关变量；去博物馆的女性比男性多，所以性别也可能是一个重要的变量，但是可能没有年龄那么重要。性别差异除了在最年轻组内差异大些外，在其他组内并没有多大。仅仅知道一个变量就能帮助我们解释博物馆参观率；知道性别和年龄的交互影响能对我们的分析提供更深入的信息。

在数据分析中引入第三个变量（有时也叫做控制变量），能更深入地揭示原来的二元关系。例如，起初的交叉表可能仅

仅是呈现年龄和博物馆的参观率情况，加上性别这个第三个变量之后，我们得到了关于年龄与参观率在不同性别间存在差异的信息。这个发现表明，年龄是个很重要的有关因素，因为它在第三个变量的每个组别中都有效应。

练习框 2.3

这是《普通社会调查》中的一个 SPSS 结果输出表。你如何解读这张二元统计表？关于应答者对在公立学校中开展性教育的态度，这张表格告诉了你些什么？

公立学校中的性教育*地区交叉表

			地区				
			东北部	中西部	南部	西部	总计
公立学校的性教育	赞成	人数	80	126	124	79	409
		占该地区的百分数	89.9%	89.4%	75.2%	83.2%	83.5%
	反对	人数	9	15	41	16	81
		占该地区的百分数	10.1%	10.6%	24.8%	16.8%	16.5%
总计		人数	89	141	165	95	490
		占该地区的百分数	100.0%	100.0%	100.0%	100.0%	100.0%

问　题

1. 哪个是因变量？哪个是自变量？它们的测量水平分别是什么？
2. 拟检验的假设是什么？
3. 为什么列联表适合这里的情况？
4. 用文字描述这个表格告诉了我们些什么。例如：第二行第三列中的 24.8 是什么意思？请用文字表达。你可以这样开始：24.8% 的……
5. 用日常生活语言对这里的 SPSS 结果输出作个结论。
6. 你能否得出结论说，这里有或者没有地区差异？还有什么可以帮助你作出结论？
7. 对于这里的结果，可以说得过去的解释是什么？有没有什么理论或先前的研究指引你解释这些发现？

交叉表总结

表格是一种展示两个变量间关系的有效方法,也是一种重要方法。一般情况下,表格提供了两个离散变量间关系的描述信息,这些变量的典型测量水平是定类水平或定序水平。它们经常被用作人口统计学变量,以提供这些变量不同类别间差异的信息。

通过引入其他变量可以编制出更复杂的表格来,不过,如果子表格太多,解读起来就比较困难。第三个变量(即控制变量)的引入,有助于我们确定原来的关系在不同的情况下是否仍然存在,或者原来的关系是否是伪关系。例如,在表2.6中就可以引入第三个变量——年龄。年龄可分为"30岁以下"、"30~50岁"、"50岁以上"几个组别。这样,就可以衍生出三个年龄组对RAP音乐意见和种族关系的统计表。通过评价原来表中发现的关系是否在不同的年龄组中仍然存在,你就可以确定年龄在解释或推翻原来的关系中所起的作用,并表明年龄在解释对RAP音乐的态度中是如何与种族因素一起相互作用的。这个方法有时被称为深入分析,或者通过对第三个(或更多的)变量的控制来分析。

如果表中的数据有某种规则,看一眼你就可以初步确定变量之间的关系。不过,要更准确地确定这种关系,还需要等到数据分析之后。下一章,我们将展示交叉表以及有关统计量的解释。

第3章　解释相互关系

统计表格做好之后,就可以对数据中的规律或者结果进行直观分析了,再之后,就是通过数学计算来对结果作定量分析。有时候,在我们的主观视觉诊断之外还需要更为客观的分析。无论是理解表格的意义还是解释统计量,都需要通过实践来学习。这一章,我们就如何理解交叉表中的常用统计量和其他二元关系中的重要统计量,拟提出一些原则性的指导意见。这里的二元关系具体指卡方检验和皮尔逊积矩相关。

定 义

■ **卡方检验(χ^2)**:卡方检验是这么一种检验方法,它通过评估变量在每个类别上的期望频数和实际观察到的频数之间的差异是否显著来度量变量之间的关系。卡方检验被认为是一种非参数检验方法。也就是说,卡方检验并不假定数据是正态分布的,而数据的正态分布假定是通过样本统计量推断总体参数时所要求的。

■ **皮尔逊积矩相关系数** r:积矩相关系数是用来度量两个变量是如何共同变化(协变)的,即一个变量上的变异是否和另一个变量上的变异相联系。计算得出的相关系数描述了变量间联系的强弱程度:系数越接近1.0,说明关系越强;越接近0,说明关系越弱。此外,联系的方向性用负号(-)表示,没有负号指的就是正号(+)。无负号的系数表示变量间的联系是正相关,即当一个增大(减小)时另一个也同时增大(减小);有负号的系数表示变量间的联系是负相关,即当一个增大(或减小)时另一个却同时减小(或增大)。

假 定

■ 对于变量类别是相互排斥的定类变量和定序变量,如果已经列出了变量的所有可能类别,卡方检验是一种理想的方法。

■ 皮尔逊积矩相关系数 r 则要求变量的测量水平是定距/定比。不过,即使是定类二分变量或定序变量,特别是在变量取值间的距离看似相等的情况下,很多研究者依然热衷于计算皮尔逊积矩相关系数 r。皮尔逊相关系数 r 假定两个变量间的关系是线性的。

取自大众出版物的真实数据

尽管大量的新闻报纸和杂志都列出了表格数据，但事实上却很少报道统计结果的显著性。至少理性的读者会对大众出版物列出的数据提出质疑：所报道的关系是真的具有意义，还是偶然性造成的结果？举个例子，一份来自《洛杉矶时报》的调查发现（2004 年 4 月 24 日的报道），18 ~ 29 岁的加利福尼亚成年人“最强烈地支持印第安人有权在自己的土地上拥有赌场，支持率为 15% ~ 75%。随着年龄的增大支持率有所下降，在 65 岁及以上的人群中，尽管仍然强烈支持，但强度已降至 31% ~ 62%”。但报纸并没有报道 75% 和 62% 之间的差异是否在统计学上显著。读者应该问一问，在印第安人经营赌场问题上，年龄和支持率之间的关系是否显著？卡方检验将有助于我们对类似发现进行解释和说明。

表 3.1　应答者与未应答者的背景特征差异

特　征	全样本/%	应答者/%	未应答者/%
人口统计学和家庭特征			
性别			
男	43.8	41.4	55.2***
女	56.2	58.6	44.8
学生申请时的年龄			
13 岁或以下	8.6	8.7	8.2**
14 岁	35.6	36.7	30.5
15 岁	46.1	45.8	47.2
16 岁或以上	9.7	8.8	14.1
民族/种族			
西班牙裔	56.2	56.7	53.9
黑人	30.2	29.3	34.5
白人	6.4	6.5	5.9
亚裔或土著	7.2	7.5	5.6
讲有限英语者	7.6	7.6	7.3

续表

特　征	全样本/%	应答者/%	未应答者/%
与谁一起住			
父母双亲	61.7	63.3	54.3**
母亲单亲	28.6	27.9	31.8
父亲单亲	4.6	4.2	6.6
其他家庭成员/非亲属	5.1	4.6	7.3
样本大小	1 764	1 458	306

显著性水平:*** $p=0.01$, ** $p=0.05$, * $p=0.10$。

来源:Kemple (2004)

作者以及标题撰写人都可能会在没有统计检验支持的情况下就贸然得出结论,要么说存在差异,要么说两变量之间强烈相关(本例中关于年龄和对拥有赌场的支持率之间的关系),这种现象非常普遍。一个富于批判精神的读者理应对结论的显著性提出疑问,特别是在百分率和其他发现结果的差异看上去很小的情况下。

来自官方的或者其委托机构和企业的研究报告中,报道结果统计检验显著性的可能性要大一些。例如,表3.1提供的数据就是来自MDRC这个非营利性研究机构的。在这份写于2004年3月的高中"职业学院"的材料中,研究人员所关注的是,和没有参加这些特别课程的学生相比,参加这一特别项目的学生,在职业选择和继续深造方面是否有所改善。数据是关于学生高中毕业四年之后的情况。"职业学院"是一种小型学习共同体,在这个共同体中,理论知识、技术教育和地方雇主的实际工作经验结合在一起。

所有的研究人员都必须应对调查问卷未应答和未完成这两种问题。对于发现的结果,一种可能的解释是:应答者和未应答者之间存在某种差异,于是对结果产生一定的扭曲。MDRC的研究报告反映出,毕业四年之后,对问卷调查作出应答和未作出应答的学生之间的差异是否显著。调查对象既包括参加过职业学院者,也包括没有参加过职业学院者。

找出变量:表 3.1 中用来比较两组调查对象情况的变量有性别、年龄、种族/民族、讲有限英语者的百分比以及与谁一起住。其中,性别为二分定名量数;年龄为定序量数,分四个级别;种族/民族也为定名量数,分四个类别;讲有限英语者的百分比为定距/定比量数;与谁一起生活又是定名量数,分四个类别。表 3.1 是五个独立交叉表的合并,包含五个卡方显著性检验,虽然表中没有列出每个检验的具体卡方值。

解读图表:卡方检验适用于检测群体中应答者和未应答者之间是否存在不同,即他们的年龄、种族、语言或家境与他们对调查问题的应答情况之间是否存在某种关系。

检验结果表明,五个变量中的三个在应答者和未应答者之间存在显著的不同。与男性相比,女性更具参与性。在全体应答者中,女性占 58.6%;而在未应答者中,女性仅占 44.8%。如果比对的两组相同,我们也期望它们的百分比相当。

进一步讲,未应答者的年龄更大,且较少来自双亲健全的家庭。举个例子,我们注意到,63.3% 的应答者同双亲一起居住,而未应答者仅有 54.3% 同双亲一起居住。在 16 及 16 岁以上人群中,应答者为 8.8%,未应答者为 14.1%。

不过,对于这两个量度,卡方检验结果显示,差异并不显著。在讲有限英语和种族/民族两个维度上,参加过职业学院者和没有参加过职业学院者的情况在统计学上非常接近。

得出结论,解释结果:样本数据显示,在应答者中,年龄更小、身为女性和来自双亲健全家庭者的比例更高,但在未应答者中,这种背景的调查对象的比例则更小。这一差异对结果可能有影响,因而使得把研究结果推广到原始样本遇到一定困难。以后进一步的数据分析可能告诉我们,样本所反映出的明显差异可能与参加“职业学院”项目与否无关,而是由于调查对象的其他特征造成的。

调查对象为什么对问卷调查和跟踪调查没有应答? 可能有很多理由。为什么调查对象中的年轻女性比其他人更多地作了应答? 这一点,根据这个样本还不能确定。不过,通过与其他调查结果相比较,我们可能得到一些关于这些问题答案的暗示。

练习框 3.1

正如表3.1所示,为了使两组间的对比公平,研究人员的典型做法是,分析调查对象的人口统计学特征,看不同人口统计学特征的子群体在比对变量上的差异是否显著。理想的情况下,每组调查对象的构成相似,以便控制那些会对输出结果产生影响的特征。在这份由 MDRC 完成的关于高中"职业学院"的报告中,研究者想落实,关于参加了"职业学院"课程的学生和没有参加该课程的学生之间的比较,是否受到两个样本其他重要特征之间差异的影响。

仔细阅读下列表格,确定在对调查作出应答的调查对象(高中毕业四年后)中,参加了"职业学院"课程的学生和没有参加该课程的学生的情况是否相似,以及他们分别与总体情况的异同。例如,在全体调查应答者中,男性占78.1%,参加了"职业学院"课程的男性占77.3%,没有参加"职业学院"课程的男性占79.1%。

高中毕业四年后学生状况跟踪调查回收率(全样本及选择的子样本)

子样本	样本容量	全体/%	参加者组/%	未参加者组/%
全样本	1 764	82.7	83.3	81.9
性别				
男	773	78.1	77.3	79.1
女	991	86.2	88.1	83.9*
民族/种族				
西班牙裔	972	83.1	84.0	82.1
黑人	523	79.9	82.2	77.1
白人	111	83.8	79.0	88.9
亚裔或土著	124	86.3	84.5	88.7
文化程度期望				
无所谓是否大学毕业	614	82.6	81.8	83.5
大学毕业	671	82.4	83.8	80.6
大学毕业后继续深造	448	82.8	84.8	80.7

资料来源:MDRC 计算结果,数据取自《高中毕业四年后"职业学院"效应跟踪评价调查数据库》。

注:采用卡方检验诊断参加者和没有参加者之间应答率的显著性。显著性水平标记为 *** $p=0.01$, ** $p=0.05$, * $p=0.10$。

来源:Kemple (2004)

问　题

1. 都有哪些变量？它们的测量水平（定名、定序、定距、定比）分别是什么？
2. 为什么用了卡方检验？
3. 用文字描述这个表格告诉了我们些什么。例如：第二行第一列“男”之后的 773 指的是什么？请用文字表述最后一个数字 80.7 的意思。你可以这样开始：80.7%的……
4. 就表格中的内容得出结论。一个星号（ * ）告诉我们什么？你得出的总体统计学结论是什么？
5. 用日常生活语言表述研究发现。研究者能否得出结论说，这两组对象相似或不相似？研究者为什么应该或者不应该对两组之间几乎不存在显著性差异而感到沮丧？

取自学术论文的真实数据

在学术论文中，很少有只有数据表格而没有显著性检验的情况。学术论文的常见目的就是阐明两个或者两个以上变量间的关系，而卡方检验和皮尔逊相关系数（r）就是常用的数据统计分析的手段。

皮尔逊相关系数 r

Alfred Adler 是一名著名的心理学家，他对个体的心理出生顺序及对其生活的影响表现出浓厚的研究兴趣。他认为，由于成长中的不同经历，人们会对自己在家庭中的角色作出不同的定位。这种自我角色意识将决定人们生活中的行为和反映方式。为了检验这一想法，Gfroerer 等（2003）通过对若干标准测量结果的描述统计，研究了 125 名女性大学生的情况。然后，

就这些女大学生在不同标准测量结果间的相关进行了分析,以检验她们意识到的对父母及兄弟姐妹的反映方式(心理出生顺序)是否与她们自己的生活方式相关(见表3.2)。

表3.2 生活方式特征与心理出生顺序间的相关分析

基础—A量表	头生	中间生	尾生	独生
所属/社交兴趣	0.138	-0.372**	0.307**	0.005
遵从	0.112	-0.339**	0.018	-0.220*
责任	0.311**	0.054	0.106	0.071
欲得到认可	0.322**	-0.132	0.254*	0.162
谨慎	-0.178	0.463**	-0.301**	0.270**
艰难	-0.197	0.380**	-0.193	0.131
权利	0.094	-0.225*	0.476**	0.228*
被众人所爱	0.179	-0.007	0.136	0.139
追求完美	0.439**	-0.367**	0.295*	0.049
顺利	0.085	-0.470**	0.265*	-0.105

注: $*p<0.05$, $**p<0.01$。心理出生顺序用《怀特—坎贝尔心理顺序调查表》测定。

来源:Gfroerer et al. (2003) [研究检索索引号:9974951]

找出变量:根据他们的论文,该研究采用《基础—A量表》来测定被调查者的生活方式特质,如属于哪类群体、遵从规则、承担领导责任、欲得到成功认可,以及小心谨慎地待人。另外,还用了五个量度来刻画生活背景,作为她们生活方式特质的可能解释:她们童年是否特别艰难或者很顺利,她们想得到多少他人的注意,她们在多大程度上想取悦他人以得到众人的喜爱,她们是否竭力追求完美且避免差错。

心理出生顺序关注的是,人们如何看待自己和兄弟姐妹之间的互动方式,这也许和他们的实际出生顺序没有关系。《怀特—坎贝尔心理顺序调查表》共有40个项目,其功能是测定学生心目中的自我家庭定位或角色,测定结果的分数反映了受测的生活方式特征。因此,心理出生顺序是自变量,测量水平为定距/定比水平。

同样,用各《基础—A量表》测得的分数也是定距/定比水平的量数,并假定依赖于心理出生顺序,因此是因变量。对于

度量定距/定比水平量数间关系的强度和方向,用皮尔逊相关 r 分析最为合适。

解读图表:相关性有两部分内容:关系的强度和关系的方向。关系的强度用相关系数的数值大小表示,取值范围为 0 ~ 1.0(1.0 为最强);关系的方向用正、负号表示,其中正号不标出来,负号表示两个变量的变化方向相反(即一个变量随着另一个变量的增加而减小)。关系的强度不同于显著性水平,后者受样本大小的影响,但前者不受。例如,弱相关在大样本的情况下很可能被判定为相关显著,而强相关在小样本情况下有可能被判定为相关不显著。因此,相关系数的数值是关系强度的最好指标。大体上讲,系数低于 0.2 或 0.3 表示弱相关,系数为 0.3 ~ 0.6 表示中相关,系数为 0.6 ~ 0.8 表示强相关,系数在 0.8 以上表示高强相关。

例如,在表 3.2 中我们可以看到,对于女学生,"头生"与"追求完美"之间呈中相关(相关系数为 0.439),但显著性水平却为 0.01。就相关关系而言,强度远比显著性水平更为重要。头生心理定位的女生表现出更"追求完美",更愿意担任"领导",更主动寻求认可或承认。

我们还可以注意到,对于"中间生"的女学生,在七对显著性相关的关系中,五对变量的变化趋势相反,这就意味着"中间生"的女生感受到归属感或者乐意与他人相处并且保持和睦的可能性较小。此外,她们还感觉到自身的成长经历较为艰难。反向关系向我们表明,被调查者在一个变量上的得分越高,在另一个变量上的得分则趋向于越低;正向相关关系则表明,被调查者在一个变量上的得分越高,在另一个变量上的得分也趋向于越高,在一个变量上的得分越低,在另一个变量上的得分也趋向于越低。

得出结论,解释结果:在表 3.2 中,皮尔逊相关 r 很少呈强相关关系,只有四个相关系数在 0.40 以上。不过,这种程度的关系与 Gfroerer 等(2003)中讨论的 Adlerian 理论的期望相一致。据此我们便可以得出这样的结论:心理出生顺序和生活方

式存在某种联系。

还有一点非常重要,需要牢记:这些统计结果都是关于被调查者的自我心理感觉的,并不代表她们的真实行为。相关分析仅向我们展现了变量之间的某种联系,但并没有告诉我们哪个变量是因,哪个变量是果。考虑到两项测定是同时进行的,很难排除有关生活方式的自我察觉会影响她们对自己成长历程中的家庭角色的解读。该研究所提供的是一系列有效关系,而不是因果解释。

练习框 3.2

对小学教师进行了一项关于师生互动情况的调查,目的是了解师生关系的质量和教师对以下问题的自我察觉之间的关系:压力、对行为困难学生的负面情感/发火以及应对困难情景的自我效能。

教师压力用压力量表测定。该量表由3个5级或7级项目构成,高等级代表压力大。例如有这样一个项目:"因为不得不处理课堂上的行为表现问题,我已考虑离开教师这个行当。"在其他3个项目(如"当学生不听指挥时我会很生气")上获得高分导致负面情感上的高分。6个其他项目(如"我具有某些正面特征,在我遇到学生问题时很管用")构成"自我效能"量表。研究中还让教师对自己和班上同学关系的百分数作出估计,关系从"很好"到"很不好"。"很好"和"好"合并为"好关系"分,"很不好"和"不好" 合并为"差关系"分。根据对113名教师的调查结果分析,得到以下皮尔逊相关系数 r 矩阵:

变量间的相关系数

变　量	1	2	3	4	5
1. 教师压力	1.00				
2. 负面情感	0.48**	1.00			
3. 自我效能	-0.45**	-0.50**	1.00		
4. 好关系	-0.14	-0.02	0.02	1.00	
5. 差关系	0.31**	0.21*	-0.11	-0.57**	1.00

注：* $p<0.05$；** $p<0.01$

来源：Yoon (2002)　[研究检索索引号：8768302]

问　题

1. 哪些是自变量？哪些是因变量？它们的测量水平分别是什么？
2. 所检验的假设是什么？
3. 为什么皮尔逊相关系数 r 是合适的？
4. 用文字描述这个表格所显示的统计量。例如：第二行第一列的相关系数 0.48 告诉我们什么？解释有些相关系数前面的负号。为什么有些相关系数的值是 1.00？
5. 用日常生活语言表述各相关系数。
6. 你能否很踏实地得出结论说，该研究所提出的理论或所定的目标已经得到证据的支持？或者尚未得到证据的支持？换句话说，教师在三个自变量上的自我察觉是否和他们如何刻画自己和学生之间的正面关系和负面关系有联系？
7. 这些结果的可能解释是什么？有没有理论或者以前的研究可以引导你解读该研究中的发现？

显著性的卡方检验

如表 3.3 所示，对于定名变量和定序变量，卡方检验是理

想的检验方法。在该研究中,研究人员所关注的是60名年长的患有慢性关节炎的女性是如何自我照料以达到独立生活的。自我照料被定义为旨在提升自我健康状况的活动能力。研究问题的焦点是不同环境下自我照料效果间的差异。不同环境指的是"长期照料机构"、"辅助生活安排"或者"回到社区家中"。与"辅助生活安排"相比,"长期照料机构"的环境受控更严,提供了更多的定期照料和援助。"辅助生活安排"只是在提出请求时才提供援助。

找出变量:生活环境是定序水平的自变量,其取值为三个定序类别:在社区中生活、有援助的生活、长期照料。研究人员根据每个类别提供援助的多寡,把这些取值定义为定序变量的取值,其中在本社区生活的受到的援助最少,因而最为独立;长期照料环境控制最严(受到的援助最多),因而最不独立。共有11个因变量,代表了自我身体照料和锻炼的不同种类,这些都是二分定名变量,代码为"报道有"或"报道没有"。从本质上讲,11个卡方显著性检验交叉表都是这样生成的。

解读图表和统计量:表中并没有报道具体的卡方值,只是对达到统计显著性水平的用星号作了标记。例如,生活环境中受到援助者有63%使用了拐杖,相比之下,接受长期照料者有8%使用了拐杖,本社区生活的人中有37%用了拐杖。这些差异即使在0.001水平上,也具有统计显著性。0.001水平的意思是,凭偶然因素获得该大小卡方值的概率是千分之一。

生活在社区自己家中的人,使用助步架和轮椅的情况要明显地少于那些生活在受援助和长期照料环境中的人;受长期照料的女性显著地缺乏有规律的行走锻炼,而生活在社区自己家中的女性有着更多的水中运动;其他方面的差别在统计学上则不显著。

表 3.3　不同环境下女性各种自我照料行为及锻炼的百分比

	女性的百分比		
	社区家中	受援助	长期照料
	%(*n*)	%(*n*)	%(*n*)
身体自我照料方式			
用夹板	37(7)	25(4)	12(3)
热敷	58(11)	81(13)	64(16)
冷敷	21(4)	19(3)	8(2)
按摩	63(12)	75(12)	60(15)
用拐杖	37(7)	63(10)***	8(2)
用助步架	16(3)*	56(9)	60(15)
用轮椅或车	16(3)*	44(7)	68(17)
用其他援助	47(9)	31(5)	16(4)
锻炼方式			
走路	63(12)	75(12)	4(1)***
水中运动	16(3)*	13(2)	0
一般性锻炼	89(17)	88(14)	68(17)

注：* $p<0.05$，** $p<0.01$，*** $p<0.001$。

来源：Baird, Schmeiser ,and Yehle (2003)　[研究检索索引号:10953598]

得出结论，解释结果：在所调查的 11 种行为中，有 5 种表现出基于环境的显著性差异。结合文中的其他数据，该文的作者们得出结论：生活环境对于老年女性为了保持自己的健康所采用的自我照料方法以及采用的程度，表现出了一定的影响。长期照料机构本来就是针对那些自我照料有较多困难的人设置的，因此，这些发现和直觉一致，较容易理解和接受。本研究所揭示的是：各种具体的自我照料行为在三种不同的生活环境中是各不相同的。

练习框3.3

该研究的对象是4 176 名脊椎骨外伤患者,研究人员想了解不同民族/种族之间,在受伤的原因(病因学)和严重程度上是否存在差异。这组研究对象中,非裔美国人占研究所涉及的少数民族的90%。研究中,下体瘫痪指的是下半身体瘫痪,四肢瘫痪指的是两个上肢和两个下肢都瘫痪。

少数民族和非少数民族受伤特性差异

变　量	少数民族/% (N=1 468)	非少数民族/% (N=2 708)	χ^2	p
受伤严重程度			7.4	<0.01
下体瘫痪	52.0	47.6		
四肢瘫痪	48.0	52.4		
受伤原因			953.8	<0.001
车祸	25.2	48.2		
暴力	49.4	8.5		
运动	2.2	12.0		
摔伤	17.0	22.8		
其他	6.3	8.5		

来源:Burnett, et al. (2002)　　[研究检索索引号:7172028]

问　题

1. 哪些是自变量?哪些是因变量?它们的测量水平分别是什么?
2. 所检验的假设是什么?
3. 为什么这里用了卡方检验?解释卡方检验。
4. 用文字描述这个表格所展示的内容。例如:用文字说说"车祸"之后的数字25.2是什么意思?你可以这么开始:25.2%的……
5. 用日常生活语言对该研究作出结论。
6. 你能否得出结论说,少数民族和非少数民族在受伤原因和受伤程度上有或者没有差异?
7. 这些结果的可能解释是什么?有没有理论或者以前的研究可以引导你解读该研究中的发现?

SPSS 输出

卡方分析

是否应该给 14～16 岁年龄段的青少年提供避孕用品？把《普通社会调查》结果中对于这个问题的回答情况按照四个不同年龄分段，汇总在表 3.4 这张交叉表中。

找出变量：因为关于提供避孕用品的态度是无法“造成”年龄的，因此，态度是因变量，年龄是自变量。从表 3.4 中可以看出，各年龄段的百分比之和正好等于 100%，核验了这些值是自变量的取值。态度采用利克特氏态度计测定。尽管这一态度测量工具是定序型态标①，但看起来似乎是等距的。注意，交叉表中的年龄段是定序的，而不是定距的，因为各年龄段的年龄跨度不一，所以，该研究中不能把年龄段当做似乎是等距的。由于是定序量数，考量它们之间的关系，卡方检验是适合的。

解读图表和统计量：看上去，年轻应答者（18～29 岁）显然非常同意对青少年提供避孕用品。事实上，与 50 岁以上年龄段相比，年轻应答者中表示非常同意的比例大约高出一倍（40.1% 对 19.5%）。可是，这样的差异在统计学上显著吗？

从表 3.4 中可以看出，皮尔逊卡方值为 61.4。卡方值不像相关系数，它的值本身并不表示任何的关系强度或方向。因此，有必要查看一下相应的显著性水平，表中用“渐近显著性（双尾）”表示。“渐近”说明检验所假定的分布是渐近分布，“双尾”说明检验是双尾检验。表中的自由度（*df*）等于频数表的列数减 1 乘以行数减 1。数字 0.000 的意思是，对于一个给定单元格数的频数表，单凭机遇，该卡方值发生的概率（可能性）肯定小于 0.001。具体而言，对于一个自由度为 9 的频数表，单凭偶然性，获得卡方值 61.4 的可能性要小于千分之一。

---

① 态标即测量态度时所用的标尺。

表 3.4

交叉表							
			年龄段				
			18～29	30～39	40～49	50+	总计
对14～16岁青少年提供避孕用品	坚决同意	人数	79	70	45	70	264
		占段内%	40.1%	30.7%	23.7%	19.5%	271%
	同意	人数	54	85	69	91	299
		占段内%	27.4%	37.3%	36.3%	25.3%	30.7%
	不同意	人数	41	38	49	102	230
		占段内%	20.8%	16.7%	25.8%	28.4%	23.6%
	坚决不同意	人数	23	35	27	96	181
		占段内%	11.7%	15.4%	14.2%	26.7%	18.6%
总计		人数	197	228	190	359	974
		占段内%	100.0%	100.0%	100.0%	100.0%	100.0%

卡方检验			
	值	自由度	显著性(双尾)
皮尔逊卡方	61.408[a]	9	0.000
似然比	60.387	9	0.000
线性—线性联系	45.043	1	0.000
有效案例数	974		

注:a 没有一个单元格中的期望频数小于5,最小期望频数为35.31。

卡方检验不是强度检验。它只能够告诉我们变量之间存在某种联系。要了解变量间的联系强度，就必须计算相关系数。针对变量的不同测量水平，要选用不同的相关关系计算方法。例如，若要计算两定序变量间的相关系数，就得计算肯得尔 τ 或者 γ 这两个统计量；若两个变量是定名变量，那么就要计算统计量 λ。

得出结论，解释结果：尽管单凭偶然性得到该研究中的发现是千分之一(0.001)的事件，我们仍能够据此得到两变量之间存在显著性关系的统计学结论。对为青少年提供避孕用品的态度，取决于应答者的年龄。切记，我们是对一群人作结论的，而不是对某一个人作结论的。如果据此便得出结论说，某20 岁的年轻人强烈支持为青少年提供避孕用品，这个结论很可能是错误的。表 3.4 告诉我们，和其他年龄段的人相比，年轻人更可能同意为青少年提供避孕用品，而不是年轻人中的每一个人都同意。

为什么年龄不同会造成如此大的差异呢？表中的数据并没有直接提供答案，不过，这组数据中的其他变量可以用来进一步探究造成差异的原因。例如，与宗教信仰、教育程度或者政治观点是否有关？与 50 岁及以上组相比，年轻组是否较少信仰宗教？教育程度更高？政治观点更加自由？或许因为他们与目标人群的年龄更接近？也许年轻一代成长的文化环境中更加性开放？在这些点上，文献回顾很可能帮助研究人员得出更加合适的结论。然而，无论如何，在没有进一步的数据分析的情况下，我们都不可作出以上结论。为什么不同年龄组间会有差别，表 3.4 中的数据或统计结果是不能解释的。

练习框 3.4

受过大学教育(大学毕业)的人和没有受过大学教育(无大学文凭)的人关于爵士乐的感觉不一样吗？

爵士乐(3) * 大学文凭交叉表

			大学文凭		
			无大学文凭	有大学文凭	总计
爵士乐	喜欢(3)	人数	539	194	733
		占组内%	48.6%	56.7%	50.5%
	喜欢不喜欢交加	人数	270	101	371
		占组内%	24.3%	29.5%	25.6%
	不喜欢	人数	300	47	347
		占组内%	27.1%	13.7%	23.9%
总计		人数	1 109	342	1 451
		占组内%	100.0%	100.0%	100.0%

卡方检验

	值	自由度	显著性(双尾)
皮尔逊卡方	25.523[a]	2	0.000
似然比	27.860	2	0.000
线性—线性联系	17.813	1	0.000
有效个案数	1 451		

注:a 没有一个单元格中的期望频数小于5,最小期望频数为81.79。

问　题

1. 哪些是自变量?哪些是因变量?它们的测量水平分别是什么?
2. 所检验的假设是什么?
3. 为什么这里用卡方检验是合适的?可以用皮尔逊 r 相关分析吗?解释卡方检验。
4. 用文字描述这个表格所展示的内容。例如:用文字说说数据部分第三行第一列的数字27.1%是什么意思?你可以这么开始:27.1%的……
5. 用日常生活语言对该研究作个结论。
6. 你能否得出结论说这里有或者没有差异?
7. 这些结果的可能解释是什么?有没有理论或者以前的研究可以引导你解读该研究中的发现?

皮尔逊相关 r

许多高中生、大学招生人员以及其他关注标准化测试作用的人，很想知道学生大学课程成绩和 SAT 成绩之间的关系。表 3.5 是关于近 2 000 名大学生的皮尔逊相关 r 分析结果。

找出变量：分析中用了三个变量：SAT 总分、高中平均级点分、大学累积平均级点分。每一个变量都属于定距/定比量数，其中 SAT 的最低可能分数是 400 分，最高可能分数是 1 600，本研究中的最低平均级点分是 2.0，最高平均级点分是 4.0。对于这种定距/定比水平的变量，研究它们之间关系的强度及其方向，皮尔逊 r 系数是合适的统计方法。

表 3.5 SAT 总分与平均级点分的皮尔逊相关分析结果

		SAT 总分	高中平均级点分	大学累积平均级点分
SAT 总分	皮尔逊相关系数	1	0.184**	0.161**
	显著性(双尾)	.	0.000	0.000
	人数	1 870	1 805	1 637
高中平均级点分	皮尔逊相关系数	0.184**	1	0.347**
	显著性(双尾)	0.000	.	0.000
	人数	1 805	1 919	1 681
大学累积平均级点分	皮尔逊相关系数	0.161**	0.347**	1
	显著性(双尾)	0.000	0.000	.
	人数	1 637	1 681	1 752

注：** 在 0.01 显著性水平上相关(双尾)。

现在，我们来看一看表 3.5 中的各项内容。这是一份 SPSS 输出表，我们首先会注意到，表的主对角线之上的相关系数与对应的主对角线之下的相关系数相同。对角线上的相关系数是每个变量与自身的相关，自然是完美相关，系数均为 1。穿过对角线上的所有 1 画一条直线，我们会发现，对角线之上的数值与对角线之下的数值相同。这就是说，“高中平均级点分”与“SAT 总分”之间的相关（相关系数为 0.184）就是“SAT 总分”

与“高中平均级点分”之间的相关。无论两个变量中的哪一个是自变量或因变量,量度它们之间关系的皮尔逊 r 都一样。可见,皮尔逊 r 是一个对称性统计量,并不假定变量之间的先后顺序或因果关系。在学术论文中不能把这张表原封不动地搬上去,而应该把重复的内容删掉。

每一个单元格都列出了两个变量参与相关分析的有效被调查者人数。例如,位于左上角的第一个单元格告诉我们,用于分析的这组数据是来自 1 870 名学生的 SAT 分数,该单元格之下的那个单元格告诉我们,我们同时有 SAT 分数和高中平均级点分的学生只有 1 805 人。

相关系数的取值范围是 0 ~ 1.0,SPSS 输出结果把它放在相应单元格的顶部,即两个变量的交叉点上。相关系数前可能带一个“ - ”号,以表示两个变量之间的关系是反向的,具体意思取决于变量的赋值规则。在这项研究中,每个相关系数都是正值。最后要说的是,每个单元格中的数据都呈现了皮尔逊相关的显著性水平,即单凭偶然性获得该系数值的概率。虽然显著性水平可能小于 0.01,但是,SPSS 统计软件仅在输出表的下面用星号(*)标出 0.01 显著性水平。

解读图表和统计量:从表 3.5 中可以看出,高中平均级点分和大学累积平均级点分之间的关系最强,相关系数为 0.347。这是一种中等强度的正相关关系,表明:高中平均级点分高的学生在大学中的累积平均级点分也可能高。另外一种解释皮尔逊 r 相关的方法是给相关系数平方。这个结果可以让我们透过一个变量的变异方差看另外一个变量的变异方差。以这个相关系数为例,$0.347^2 = 0.120$,即 12%。这告诉我们,因变量上 12% 的变异方差可以用自变量来解释。有时候把这个值称作误差降低的比例,简称误降比。意思是说,如果知道被试样本在高中平均级点分上的变异方差,我们就可以把预测或解释这组被试在大学累积平均级点分上的变异方差的误差方差降低 12%。显然,大学累积平均级点分变异方差的 88% 要由其他变量来解释,比如说学习时间、智力水平、学习动机、课程

难度和教师等。可见,相关系数为 0.347,说明变量间的关系不是很强。

皮尔逊 r 相关法的其中一个局限是:当样本容量足够大时,很小的相关系数也可能具有统计显著性。在该例子中,对于 1 850 名学生,高中平均级点分和 SAT 总分之间的相关系数为 0.184。这样的微弱相关,其发生的概率为 0.000,意即 $p < 0.001$。对于相关研究,相关系数的数值本身比显著性更为重要,因为它是变量间关系强弱程度的量度。就本例而言,SAT 总分分别与高中平均级点分和大学累积平均级点分在统计学上显著相关,但它们之间的关系并不很强。

得出结论,解释结果:表 3.5 所呈现的相关分析结果告诉我们,作为预测大学累积平均级点分的指标,高中平均级点分比 SAT 总分更好。尽管 SAT 成绩和平均级点分相关(SAT 总分较高的学生,他们中学和大学级点分也倾向于较高),但这种相关通常是微弱相关。对于解释或预测大学级点分,其他变量可能比 SAT 分数更好。

SAT 分数之所以未能较好地解释大学级点成绩,原因之一可能是这组大学生被调查者样本自身的变异小于大学生群体的应有变异。大学录生时,只录取 SAT 成绩和高中平均级点分较高的学生,这样便减短了学生的变异范围。如果选取的被调查者样本是取自变异范围更大的高中生,其中包括那些没有被大学录取的高中生,高中平均级点分与 SAT 总成绩之间的相关系数可能会高一些。不过,一些研究者觉得,SAT 成绩并不能很好地预测学生在大学中的学业表现。

练习框 3.5

初婚年龄与所成长家庭中的兄弟姐妹人数以及他们目前所养育孩子的数量有没有关系?下表是取自《普通社会调查》结果的一组数据。

相关分析结果

		初婚年龄	兄弟姐妹数	孩子数量
初婚年龄	皮尔逊相关系数	1	-0.006	-0.259**
	显著性(双尾)	.	0.831	0.000
	人数	1 202	1 199	1 199
	皮尔逊相关系数	-0.006	1	0.202**
	显著性(双尾)	0.831	.	0.000
	人数	1 199	1 495	1 491
	皮尔逊相关系数	-0.259**	0.202**	1
	显著性(双尾)	0.000	0.000	.
	人数	1 199	1 491	1 495

注：** 在0.01水平上显著(双尾)。

问　题

1. 哪些是自变量？哪些是因变量？它们的测量水平分别是什么？
2. 所检验的假设是什么？
3. 为什么这里用皮尔逊 r 相关分析？解释表中的相关分析结果。
4. 用文字描述这个表格所展示的内容。例如：负相关系数(-0.259)是什么意思？如何计算 r^2？(用文字说说) r^2 的意思是什么？
5. 用日常生活语言对该研究作个结论。
6. 你能否得出结论说这里的关系显著或者不显著？
7. 这些结果的可能解释是什么？有没有理论或者以前的研究可以引导你解读该研究中的发现？

卡方检验与相关分析小结

在阅读包含定名变量和定序变量的表格时，应该在统计学理论的指导下对结果作出解释，这一点非常重要。虽然数据的量值本身可以直观地告诉我们一些信息，但是在评价变量间的关系时，我们通常还需要一些统计量的援助。

卡方检验为我们提供的信息，能帮助我们从数学的角度判断每个频数单元格中观察到的频数与我们所期望的结果是否不同。卡方检验能告诉我们，对于给定的频数单元格和样本容量，单凭偶然性，我们能得到某个卡方值的概率是多少。

卡方值并不是变量间关系强度或方向的量度，而是它们有无联系的量度。一旦单凭偶然性获得某卡方值的概率小于0.05，变量间的联系就在统计学意义上显著（除非事先设置了更苛刻的显著性水平，比如0.01或0.001）。这时我们就可以得出结论说，因变量和自变量之间有联系。

变量间关系的强弱程度只能通过相关分析来确定。如果变量都是定距/定比水平上的，是二分量数，或至少是看似定距的定序量数（如利克特氏态标上的量数），那么，我们就可以计算皮尔逊相关系数 r。系数值越是靠近0.0，表明越没有关系，系数越是靠近1.0，表明两变量有着很强的关系。不过，自变量与因变量之间的相关关系并不表示它们之间存在因果关系。

第4章 解释均值差异

除了评价变量之间是否有关系外，研究工作者常常对比较两组或多组对象之间的差异很感兴趣。经过一段时间后，同一组对象是否发生了变化？或者不同组对象之间在某些选定的标准上是否存在差异？

分析组间或组内差异的方法之一，是用统计方法对均值上的差异加以分析。本章介绍两种用来解释均值差异的基本统计方法，即 t 检验和 F 检验（ANOVA 或方差分析）。

定 义

■ **独立样本 t 检验**：独立样本 t 检验就是计算两个独立样本在同一个变量上的均值差与特定标准误（由标准差和样本容量来决定）的比值 t，然后再把这个比值与 t 值分布表对照，以确定它的显著性水平。大体上说，对于双尾检验，如果 t 值大于 2.0，差异很可能显著。

配对 t 检验的情况也很相似。由于它比对的是同一组人在两个不同变量上的平均值之间的差异，计算比值时的方法略有不同，以便把两个变量间的相关关系考虑在内。

■ **单向方差分析**（ANOVA）：数学上，方差分析就是计算 F 值，即组间（三组或三组以上）变异的方差与组内变异的方差的比值。变异方差即相对于均值的偏离，因此，比较多个均值时用方差分析。当组间差异大于组内差异时，我们便得出结论说，所比较的三组或更多组在均值上的差异是显著的。当组间差异和组内差异差不多时，F 值大约为 1.0，这时我们便得出结论说，所比较的三组或更多组在均值上的差异不显著。

假 定

■ 对于比较两个均值之间的差异，t 检验是理想的统计技术。t 检验中的“t”本来就取自英语单词 two（二）的第一个字母。t 检验要求（因）变量在定距/定比水平取值，还必须是正态分布的；自变量则在定名水平上取两个值。对于独立样本 t 检验，两个样本必须互斥，即样本中的单位不能既可以属于这个样本，也可以属于那个样本。例如，对比新生和即将毕业的学生时，任何被调查者都不能同时既是新生也是即将毕业的学生。对于配对样本 t 检验，“两个组”由同一批人组成，不过，每

人都有两个测量结果参加比较。例如,被调查者应该由前测分数和后测分数来比较。

■ 对于比较三组或更多组在均值上的差异,方差分析是最佳选择。计算方差和均值时,(因)变量的取值必须是定距/定比水平上的量数。例如,可以用方差分析技术比较四种不同宗教背景的学生的大学平均级点分差异。通常,组别变量是定名或定序水平。

取自学术论文的真实数据

一段时间内均值上的变化和其他数据资料上的变化,通常是许多民意调查、报纸图表、各级政府报告以及一些非营利性研究团体或组织的研究报告必不可少的部分。然而,在他们向普通大众公布的材料中几乎从不提供 t 检验结果、方差分析检验结果或者其他统计检验结果。因此,本章所用的例子主要取自学术论文或著述,而不是取自大众媒体。

t 检验

在最简单的情况下,一篇论文中可以只用一句话来报告 t 检验的结果。Mayo (2002)对案例法和更传统的讲授法对心理学引论课的教学效果作了比较,用课程成绩作为教学效果的量度。看一下这篇文章在其结果部分是如何表述 t 检验的结果的:

> 用从一份有 30 个题目的试卷中获得的数据,我对接受案例法教学的学生和控制组的学生的学习成绩做了独立样本 t 检验。恰如假设的那样,接受案例法教学的学生的学业成绩($M = 83.23, SD = 10.29$)与控制组学生的学业成绩($M = 76.76, SD = 12.43$)有显著差异,$t(134) = 2.34, p < 0.01$。
>
> [研究检索索引号:5655548]

以上报告告诉我们,接受案例法教学的学生的平均成绩较高,接受传统讲授法的学生的平均成绩较低,前者为 83.23 分,后者为 76.76 分,前者高出后者大约 6.5 分。而且,接受传统讲授法的学生的成绩更分散一些,标准差为 12.43;接受案例法教学的学生的成绩则分散小一些,标准差为 10.29。对于这样的样本容量,t 值等于 2.34,在 $p<0.01$ 水平上达到显著。在这项试验中,136 人(样本容量是根据自由度 134 推算出的。自由度等于案例法班人数减 1,加上控制班人数减 1,即:$136-2=134$)参加并完成了考试。作者得出结论说,与讲授法相比,案例法是一种教授普通心理学理论更为有效的方法。

在更复杂的情况下,期刊论文中可能会出现 t 检验表,就像表 4.1 那样。在表 4.1 中,比较了五、六月份出生的学生与前一年七、八月份出生的学生在阅读能力分数、拼写能力分数以及在校行为表现分数上的平均差异。该项目把研究对象锁定在小学一年级学生上,以了解班中年龄最小的学生(入学截止出生日期 6 月 30 日以前的孩子)是否比班里年龄较大的学生(出生在前一年七、八月份的学生)更易表现出行为异常问题和较差识字能力问题。

找出变量:本研究涉及三个因变量,它们分别是:字母识别分数、单词识别分数和在校行为表现分数。每个变量的取值都在定距/定比测量水平,所以适合计算均值。在测试中,要求学生指出并念出字母表中的每一个字母,还要求学生指出并念出一组单词中的每一个字母。得分就是所识别对的字母个数。为了让教师测定学生的在校行为表现,专门编制了一套由 8 个项目组成的行为量具,内容涉及“集中注意能力”、“按指令做事能力”及一般成熟水平等,用利克特氏 5 级记分,分数越高,代表在校行为表现越好。比较在两组被调查者之间进行,一组出生在特定年的五、六月份,另一组出生在前一年的七、八月份。

研究人员用了独立样本 t 检验统计技术,以诊断两组被调查者在三个因变量上的平均成绩差异是否具有统计显著性。

表 4.1　一年级学生出生月份与识字能力及行为表现

出生月份	字母识别（平均成绩）	单词识别（平均成绩）	行为表现（平均成绩）
五月/六月（第一组,$n=17$）	17.68	14.76	20.24
七月/八月（第二组,$n=20$）	22.10	21.75	32.55
显著性水平（p 值）	$p<0.008$	$p<0.002$	$p<0.000$

来源：Menet, Eakin, Stuart, and Rafferty (2000)　［研究检索索引号：3788707］

解读图表：表 4.1 虽然没有给出有关 t 检验的值,但给出每一个变量的相应显著性水平(p)。文中不仅给出了有关 t 检验的值,而且进行了讨论。从表中可以看出,学生在每个变量上的平均分的差异都达到了一般设定的统计显著性水平。但从测试结果看,在字母识别平均分数上的差异的显著性,没有在校行为表现平均分数上的差异的显著性那么高。看上去,年龄较小的那组被调查者(五、六月份出生的学生)在字母和单词识别上比年龄较大一些的那组被调查者(前一年七、八月份出生的学生)做得差一些。同样,教师给年龄较小组的行为表现打分也低于给年龄较大组。也就是说,年龄较小组的"集中注意能力"、"按指令做事能力"以及教师不管时的自律能力,平均要低于年龄较大组。

得出结论,解释结果：从分析结果可以看出,出生月份(班级中的相对年龄)与文字识别能力和成功行为表现有一定关系。一年级学生中,年龄小一些的被调查者表现得不如同班中年龄稍大一些的被调查者。

为什么会这样？单从表 4.1 中的分析结果是不能确定的。不过,该论文的文献综述部分已经表明,其他研究也指出,同班学生的年龄差别是引起他们学业差异和行为表现问题的因素之一。教师假定,班上的所有同学,无论年龄大小,其行为能力和认知能力都一样,这对小龄学生也许期望太高了一些。

t 检验范例

本研究对 2 429 名成了家的工人进行了调查，主要针对不同的轮班对工人的家庭、工人本人的情绪、体力以及时间安排等方面产生的负面影响。表 4.2 呈现了该研究的部分结果。

找出变量：因变量是“工作对家庭的负溢入”①，通过 5 个项目来测量，每个项目有 5 个得分等级，1 表示“从未”，5 表示“总会”。题项内容包括：“在过去的三个月中，你曾多少次遇到没有足够时间和家人或其他重要的人在一起的情况？”或者“在过去的三个月中，你在家时，有过多少次心情不好的情况？”这些项目都在定距/定比测量水平，所以，可以计算该 5 项目测量工具测量结果的平均分。

表 4.2 七种轮班方式对家庭生活负面溢入的比对分析

(I)轮班	(J) 轮班	均值差(I－J)	显著性	下界	上界
傍晚	白天	7.663E－02	(NS)0.962	－0.174 5	0.327 7
夜间	白天	0.242 1	(NS)0.172	－4.7270E－02	0.475 6
每天改变	白天	0.321 0*	0.001	9.814E－02	0.543 9
二分	白天	0.375 5	(NS)0.228	－0.111 1	0.862 1
灵活	白天	7 879E－03	(NS)1.000	－0.167 0	0.182 8
其他	白天	0.301 4	(NS)0.261	－0.102 2	0.705 0

注：t 检验将其中一组当作控制组与其他组进行比对，均值显著性水平为 0.001。

来源：Grosswald (2003)　[研究检索索引号：12596307]

自变量是 7 种不同的轮班方式。例如，傍晚班是从下午 4 点左右到半夜 12 点，夜间班从半夜 12 点到次日早晨 8 点。轮班方式是定名水平的测量。如果要比较不同轮班方式间的平均差异，就得用独立样本 t 检验。如果想把各种轮班方式一次性比较，那就得用单向方差分析。不过，这项研究的研究人员

① 负溢入即工作对家庭生活的负面影响。

感兴趣的却是每一种不同轮班方式和标准轮班方式(白天班)间的均值差异。

解读图表:表4.2并没有报告“负面溢入”测量结果的具体均值,而是报告了每一种轮班和白天班之间的均值差,以及相应的 t 检验结果的显著性。六对比较中,只有一对之间的均值差达到显著性水平,表中用“ * ”标注,其余五对都没有达到显著性水平,表中用NS标志。

此外,表4.2还报告了95%置信区间的上、下界值。意思是说,在随机抽取的样本中,其中有95%的样本的均值差会落在这个区间之内。用推断统计学的语言来说就是,均值之间的真正差异,很可能(有95%)就落在上、下界之内。例如,就这组样本而言,“每天改变”轮班的工人与白天班的工人之间的均值差为0.321 0,但是,就这组样本所代表的总体而言,他们之间的均值的真正差异可能在9.814E－02(这是一种数学上的简易记数法,将小数点向左移动2位就还原为日常记数法)或0.098 14和0.543 9之间。也就是说,如果从总体中随机抽取被试样本,其中有95%的均值差都在0.098 14和0.543 9之间。通常,发表的论文并不报告置信区间,但是,该文的作者却选择了报告置信区间。

得出结论,解释结果:和只上白天班的工人相比,只有那些上班时间每天都改变的工人反映,他们的工作安排对家庭、本人情绪及体力方面带来负面影响,其他轮班方式的工人的反映,与只上白天班的工人的反映并没有显著性差异。这些结果也许表明:稳定的工作日程会使家庭生活的安排更为容易,天天改变工作日程会使家庭生活的安排变得非常困难。为什么会这样呢?该研究作了进一步的说明,并对其他变量的数据进行了分析。

练习框 4.1

本研究(Bisset et al. 1999)重点比较了学生、家长以及全体教师在大学教育目标接受上的差异。总共涉及诸如个人与社会发展、科学与技术等6个方面。对于这6个方面的理解,通过34个项目进行测量,每个项目都是利克特氏5级评分项目,5分表示非常同意。使用了独立样本 t 检验技术,以评价学生和教师对6个方面教育目标接受的差异。

学生和教师对教育目标接受情况的比较

项 目	组均值		
	学生 $N=182$	教师 $N=62$	显著性
A. 科学与技术	3.948 1	4.310 5	***
B. 人文	4.051 9	4.362 9	**
C. 社会科学与行为科学	3.861 5	4.283 4	***
D. 学力与沟通技巧	4.446 9	4.817 2	***
E. 个人发展与社会发展	4.577 5	4.615 6	NS
F. 职业发展	4.540 7	4.254 8	***

注: * $p<0.01$, ** $p<0.001$, *** $p<0.000\,1$。

来源:Bisset et al. (1999) [研究检索索引号:2984074]

问 题

1. 哪些是自变量? 哪些是因变量? 它们的测量水平(定名、定序、定距/定比)分别是什么?
2. 所检验的假设是什么?
3. 为什么这里用 t 检验合适?
4. 用文字描述这个表格所展示的内容。例如:“学生”列中与目标 E 对应的数值 4.577 5 是什么意思? NS 又是什么意思? 样本容量是多少?
5. 用日常生活语言对该研究作个结论。
6. 你能否很肯定地得出结论说,学生和教师对教育目标的接受有差别或者没有差别?
7. 这些结果的可能解释是什么? 有没有理论或者以前的研究可以引导你解读该研究中的发现?

方差分析

研究中一个常见的目标就是对三个或三个以上的组平均值进行比较。方差分析可以一次性地对这类比较的结果进行显著性检验。请看表4.3。这张表显示了自我形象感和种族身份认同感之间的关系,被调查者包括347名以南非荷兰语为母语的白人,133名以英语为母语的白人以及466名南非黑人。

找出变量:自我形象感是用标准罗森伯格《自尊心量具》来测量的,每个被调查者有两个自尊心分数,一个关于负面自我形象感的,分数范围是4~16分,分数越高,代表负面自我形象感越严重;另一个是关于正面自我形象感的,分数范围是5~20分,分数越高,代表正面自我形象感越厉害。种族认同感用20个利克特氏五分式项目测量,共分为三个子量具:①种族认同感(对种族的忠诚、尊重及自豪感,分数范围是9~45分,分数越高,表示种族认同感越强);②种族身份成就感、本种族文化活动参与度(分数范围是6~30分,分数越高,表示探索越多、成就越大、参与度越深);③不确定感(对属于该种族的不舒适感、不确定感,或者对保留和保护种族身份的不舒适感、不确定感,分数范围是4~20分,分数越高,表示越愿意保护种族身份,对所属种族的不确定态度越低)。

所有五个因变量都属于定距/定比水平的量数,所以,都适合计算方差和平均值。三个种族类别是自变量,属于定名水平。由于要在三个种族之间同时比较均值和方差,因此,单向方差分析是合适的统计方法。

解读图表:从表4.3中可以看出,所有的F值都在0.01水平达到显著。换句话说,如果各种族之间没有差异,对于这么大的样本,单凭偶然性获得这么大的F值,其可能性不会超过1%(或$p<0.01$)。

例如,在负面自我形象感的16分中,以南非荷兰语为母语的白人平均得了6.66分,南非黑人平均得了9.21分,以英语为母语的白人平均得了7.35分。三者之间的平均分数差具有

表 4.3 平均分数及方差分析结果

变 量	以南非荷兰语为母语的白人		南非黑人		以英语为母语的白人			
	M	*SD*	*M*	*SD*	*M*	*SD*	*df*	*F*
自我形象感								
负面自我形象感	6.66	2.39	9.21	2.89	7.35	2.89	2,921	99.47**
正面自我形象感	17.49	2.08	15.97	3.31	17.30	2.29	2,921	32.47**
种族认同感								
种族认同感	37.92	6.45	36.71	6.97	35.01	6.67	2,922	8.52**
成就感	22.84	4.76	20.73	5.64	21.13	4.43	2,921	16.79**
不确定感	15.36	3.33	12.64	3.65	13.42	3.14	2,922	61.71**

注：* $p < 0.01$。

来源：Bornman (1999)

统计学显著性。对于这三个种族,负面自我形象感得分的标准差基本上一样大,表明组内分数的分布也大概一样。简言之,对于这三个不同种族,组间方差大于组内方差。

再看 *df* 列,这个列是自由度。在表 4.3 中,每个报告自由度的单元格包括两个自由度,第一个自由度是参加比较的组数减 1(即 2),第二个自由度是每组被调查者的人数减 1 之和,即各个组的被调查者总和减组数(即 921)。在本研究中,被调查者分为三组,所以,对于每一个 *F* 检验,第一个数字(自由度)都是 2,至于第二个数字(自由度),其值取决于具体有效问卷的数量。对于该研究,第二个数字可能高达 923(466 - 1 = 465 个黑人,347 - 1 = 346 个以南非荷兰语为母语的白人,113 - 1 = 112 个以英语为母语的白人,总数为 465 + 346 + 112 = 923)。我们要用这两个自由度来确定 *F* 值的显著性,因为 *F* 值的显著性取决于这两类样本的大小。例如,对于 1 000 名被调查者所代表的 5 个组的民族认同感进行比较时,其自由度为 4,995,一个为 8.52 的 *F* 值是不显著的。尽管 *F* 值本身并没有什么意义,只有根据它计算出的具体显著性水平才有意义,但是,*F* 值小于 1 或者稍大于 1 时,组间的平均值差异通常是不显著的。

得出结论,解释结果:从表 4.3 的数据我们可以看出,与其他种族相比,黑人的正面自我形象感最低,负面自我形象感最高,而以南非荷兰语为母语的白人则刚好相反,他们的正面自我形象感最高,负面自我形象感最低。进一步统计分析表明,在两组白人之间,自我形象感的差别并不显著。通常,在根据 *F* 值及相应的两个自由度判断组间均值差异显著时,还要进一步检验具体哪个或哪几个组间的平均值的差异达到了显著性水平。关于该研究的进一步检验,请阅读该篇文章的全文。

另一方面,以南非荷兰语为母语的白人在三种种族认同感上的得分都显著的高,而黑人在其中两种种族认同感(成就感和不确定感)上的得分则显著的低,但是和以英语为母语的白人相比,在种族认同感分量表上的差异并不显著。根据 *F* 值判断,在三个种族民族认同感上的均值差异均达到统计显著性

水平。

是什么造成了这种差异呢？关于这个问题，根据表 4.3 中的数据还不能够回答。长期的种族隔离史可能是寻找这个问题答案的好地方和好开端。令人感兴趣的是，其他种族的成员往往表现出强烈的种族认同感和正面的自我形象。为什么会出现这些差异呢？以南非荷兰语为母语的白人有着更为强烈的种族认同感和自我形象的原因非常复杂，这里不便详细讨论。感兴趣的读者，可参阅原始论文。

方差分析范例

研究者对参加社区膳食工程的 65 岁及以上老人的孤独感水平颇感兴趣，因此，通过抽样对他们的孤独感水平进行了研究，抽取的样本人数为 849 人，结果如表 4.4 所示。

找出变量：因变量是孤独感的程度，用 11 个项目测量，每一项得 0 分表示孤独，得 1 分表示不孤独，总分等于各个项目得分之和（最低分 0 分，最高分 11 分），分数越高，表示孤独感程度越低。这样的测量结果是定距/定比水平的量数，适合做方差分析。典型的项目如："我希望我有一个真正的密友"，"我很想有人陪伴"，以及"我体会到空虚"。

根据婚姻状况、有没有孩子以及有没有密友，把被调查者分为八个不同的组。自变量是定名变量。

解读图表：表 4.4 中呈现了八组不同状况的老年人的孤独感平均值、标准差和各组的人数。千万不能忘记，分数越低，表示孤独感程度越高。从表中可以看出，没有子女和朋友的孤寡老人的孤独感程度最高。饶有兴趣的是，从平均分数（5.5 分、5.67分）上看，无配偶与有配偶但没有子女和好朋友的老人几乎有同样程度的孤独感，而有配偶有子女和密友的老人是最不孤独的（11 分中得了 8.06 分）。

F 统计量表示这些均值间的差异是否达到预先设定的显著性水平。*F* 值后面两个数值中的第一个（6.56）为比值。通常，如果不同自变量在均值和方差上的差异有显著性，*F* 值一

定要大于1才有可能。括号里面的是两个自由度,第一个自由度等于自变量的组数减去1(即8-1=7),第二个自由度等于总的有效被调查者人数减去组数8(即849-8=841)。对于这么多的组数和有效总被调查者人数,根据这组数据计算所得的 F 值的概率小于0.0000(肯定小于通常的临界值 $p<0.001$)。

表4.4 不同状况老年人孤独感方差分析结果

老年人状况	*M*	*SD*	*N*
1. 无配偶、无子女、无朋友	$5.67^{4,8}$	3.26	58
2. 无配偶、无子女、有朋友	6.61^{8}	3.23	66
3. 无配偶、有子女、无朋友	$6.39^{4,8}$	3.33	169
4. 无配偶、有子女、有朋友	$7.49^{1,3}$	2.93	275
5. 有配偶、无子女、无朋友	5.50^{8}	3.81	18
6. 有配偶、无子女、有朋友	7.09	3.20	23
7. 有配偶、有子女、无朋友	6.75^{8}	3.03	101
8. 有配偶、有子女、有朋友	$8.06^{1,2,3,5,7}$	2.99	139
	$F=6.56(7,841\ df);p<0.0000$		

注:上标指差异在统计学上显著($p<0.05$),采用Tukey氏多重显著性检验对不同状况老年人的孤独感均值差异进行了逐对比较。例如,状况1的均值与状况4以及状况8的均值差异都达到预定的显著性水平。

来源:Hall-Elston and Mullins (1999)

该 F 检验结果告诉我们,8个被调查组之间的均值差异已经达到统计显著水平。由于 F 检验一次比较的对象多(两个以上),所以,在发现 F 检验结果达到统计显著水平后,很多研究者还要进一步做“事后比较”(也叫“多重检验”)分析,从而发现哪一组和哪一(几)组之间的均值差异达到显著性水平。“事后比较”分析的方法有多种,各种方法的假定不一(各种高级统计学教材对常用的“事后比较”分析方法有较详细的介绍)。本研究中,研究者用了Tukey氏多重显著性检验法,并在表的脚注中作了说明。表中平均值右上角有数字的,表明了哪

一组的均值与该组均值的差异达到约定的统计显著性水平。以第四组为例(无配偶但有孩子和朋友),该组的平均值与第一组和第三组平均值的差异显著。从表中可以看出,第八组的均值(8.06)除了与第四组和第六组的差异不显著之外,与其他五组的均值的差异都显著。

得出结论,解释结果:表4.4中的数据和统计量都表明,老年人的孤独感与有无配偶、子女以及亲密朋友有显著关系。所有组的调查结果都显示,没有朋友的老人比有朋友的老人相对要孤独一些;没有子女的老人要比有子女的老人孤独一些。然而,无配偶的老人和有配偶的老人相比,他们的孤独感分数几乎一样。进一步比较的最好方法是,将两个在两个指标上一样(只有一个指标不一样)的组进行比较。例如,把第一组与第五组或第四组与第八组进行比较,它们之间的唯一差别是有无配偶。可以看出,他们之间的分数非常接近,统计检验结果显示,他们之间的平均孤独感差异不显著。

如果我们把第三组与第四组(这两组的唯一不同是有无密友)比较,它们之间孤独感的平均值差异就是显著的:无配偶、有子女、没朋友的老年人的孤独感比无配偶、有子女、有朋友的老年人的孤独感要更高。

大多关于友谊的研究结果都支持生活中个人社会网络是重要的这一结论,不论这个网络涉及的是子女还是朋友。但是,对这批调查对象而言,配偶的影响为什么不是很大?其中的原因不易解释清楚。不过,那篇论文从其他角度所作的探索,可能会给我们解释这些发现提供一些启发。

练习框4.2

研究者就新教师是否愿意在课堂上使用新的科技手段作了调查研究。研究者假设:在一些高小和中学代课教师中,对新科技教学手段的关注可能因人格类型的不同而不同。研究中,用目前流行的MBTI调查工具测定了被调查者的人格类型。先把调查对象按人格分为八个不同的组,再在这八个组的基础上进一步概括出四种人格类型。被调查者对于使用新科技教学手段的愿意程度,是用由20个项目构成的利克特氏态度量具测定的,分数越高,表示越打算使用新科技教学手段。

使用新科技教学手段意愿的部分统计结果(根据MBTI人格类型分类)

MBTI人格类型	教师人数	均 值	标准差
直觉/感觉型	43	100.81	11.99
直觉/多思型	23	105.04	11.47
触觉/感觉型	30	91.03	13.09
触觉/多思型	63	94.76	16.11

方差分析的显著性测验

来源	*SS*	*df*	*MS*	*F*	*p*
组间	3 491.85	3	1 163.95	6.017	0.001
组内	29 983.86	155	193.44		
总计	33 475.71	158			

来源:Chambers, Hardy, Smith, and Sienty(2003)

[研究检索索引号:10954975]

问 题

1. 哪些是自变量？哪些是因变量？它们的测量水平分别是什么？
2. 所检验的假设是什么？
3. 为什么这里用方差分析是合适的？
4. 用文字描述这个表格所展示的内容。例如：F 值和 p 值分别是什么意思？样本容量是多少？均值和标准差如何解读？
5. 用日常生活语言对该研究作个结论。
6. 你能否得出结论说，不同人格的教师在使用新科技教学手段的意愿上有差别或者没有差别？还有些什么信息可以帮助你作出结论？
7. 这些结果的可能解释是什么？有没有理论或者以前的研究可以引导你解读该研究中的发现？

SPSS 输出

配对样本 t 检验

《普通社会调查》问了很多关于人们喜欢什么和不喜欢什么方面的有趣问题，其中包括一些关于音乐种类的问题。表 4.5呈现的数据是关于被调查者对百老汇音乐剧和歌剧的观点的对比结果。

找出变量：关于百老汇音乐剧和歌剧的观点，是用同样的利克特氏 5 级态标测量的，等级从“非常喜欢”（得 1 分）到“非常不喜欢”（得 5 分）。这是一个定序量标，但由于看上去是定距的，因此当作定距/定比量标来对待。这样，便可以计算该被调查者样本的平均值。

表 4.5

配对样本统计量

		均值	样本人数	标准差	标准误
第 1 对	百老汇音乐剧	2.59	1 359	1.098	0.030
	歌剧	3.48	1 359	1.134	0.031

配对样本相关分析结果

		样本人数	相关系数	显著性
第 1 对	百老汇音乐剧和歌剧	1 359	0.452	0.000

配对样本 *t* 检验结果

		配对差异							
					95% 置信区间				
		均值	标准差	标准误	下界	上界	t	df	显著性（双尾）
第 1 对	百老汇音乐剧对歌剧	−0.89	1.169	0.032	−0.95	−0.83	−28.100	1 358	0.000

解读图表:本研究调查了 1 359 名被调查者对音乐剧和歌剧的不同观点。由于我们对比的是同一组人(而不是两个独立组)对音乐剧和歌剧的不同观点,而且都是用同样的利克特氏量标测量的,因此,合适的检验方法是配对 t 检验,而不是独立样本 t 检验。

在 1 ~5 级(1——喜欢,5——不喜欢)的等级量标上,被调查者对音乐剧和歌剧的观点的平均分分别为 2.59 和3.48,标准差几乎相同。这表明两组被调查者对每一种类型音乐的观点有很相似的离差。虽然被调查者对这两种音乐都有一种复杂感情,但总的来说,比起音乐剧来,他们好像更不喜欢歌剧:被调查者对音乐剧的态度分为 2 ~3 分("喜欢"到"复杂感情"),而对歌剧的态度分则为 3 ~4 分("复杂感情"到"不喜欢")。

第二个表中所报告的皮尔逊相关系数 r 表明,被调查者对于音乐剧的态度和对于歌剧的态度之间呈中度相关,相关系数为0.452。对于这么大的样本,已达到统计显著性。尽管两者之间有一定关系,但确定系数 R^2 表明,对一种音乐形式的态度变异,只有其中的20%才能被他们对于另外一种音乐形式的态度解释。也许,被调查者喜欢与不喜欢一种音乐形式,并不总是能通过他们喜欢还是不喜欢另外一种音乐形式看出。t 检验还能为我们提供更多的信息。

这组被调查者对百老汇音乐剧和歌剧的平均态度分之间的差异为 -0.89,其对应的 t 值是 -28.1,负号只是由于求差时被减数和减数的不同而致。例如,如果被减数是歌剧,减数是音乐剧,则均值差为 0.89;如果被减数是音乐剧,减数是歌剧,则均值差为 -0.89。对于该研究这样大小的样本(自由度为 1 358),单凭偶然性获得这么大差异的概率是 0.000,报告中这个结果可以用统计显著或 $p<0.001$ 描述。本研究构造的零假设是双尾假设,即被调查者对音乐剧的态度和对歌剧的态度没有差异。

95%的置信区间告诉我们,如果抽出很多个这么大的样

本,其中95%的差异落在置信区间的上界和下界之间(见表4.5)。用更通俗的话说就是,对于这组被调查者所代表的全体被调查者,他们对这两种音乐形式的真正态度均值差,有95%的可能落在-0.95和-0.83之间。

得出结论,解释结果:分析结果表明,尽管有些"复杂情感",但对于音乐剧和歌剧态度的差异,已经达到一般设定的统计显著性水平。总的来说,比起歌剧来,被调查者更喜欢音乐剧;或者比起音乐剧来,被调查者更不喜欢歌剧。

对此,数据并没有提供解释。也许可以通过研究关于其他变量的信息来解释这一结果,例如教育水平、社会地位或者年龄等。我们都知道,比起音乐剧来讲,歌剧更难欣赏,也更难理解,而且,歌剧的媒体形象也常常是负面的。或许这些理由可以解释被调查者对音乐剧和歌剧态度之间的差异。

独立样本 t 检验

表4.6呈现了一些来自《普通社会调查》的其他数据,是关于男性与女性首婚年龄的比对数据。

找出变量:显然,年龄是一种定距/定比变量,适合计算均值,性别是二分定名变量。要对比女性的首婚平均年龄和男性的首婚平均年龄差异,独立样本 t 检验是一种恰当的统计检验方法。

解读图表:表4.6由两个分表组成,第一个分表呈现的是男性样本和女性样本的描述统计结果,包括均值、标准差等。492位男性首婚平均年龄为24.16岁,710位女性首婚的平均年龄为21.84岁。单从平均年龄上就可以看出,男性明显比女性结婚晚,t 检验的结果也证实了这一结论,即男女在平均首婚年龄上差2.32岁具有统计显著性。

下面,我们将首先看一看男性和女性的首婚年龄分布是不是相似。除非男性的首婚年龄分布和女性的首婚年龄分布大体相同,否则简单进行比较是不公平的。独立样本 t 检验的SPSS

表 4.6

组统计量

	性别	样本人数	均值	标准差	标准误
首婚年龄	男性	492	24.16	4.867	0.219
	女性	710	21.84	4.929	0.185

独立样本 t 检验结果

		Levene 氏方差齐性检验		均值差异 t 检验					95% 置信区间	
		F	Sig	t	df	Sig(双尾)	均值差	标准误	下界	上界
首婚年龄	假定方差齐性	0.342	0.559	8.066	1 200	0.000	2.32	0.288	1.756	2.885
	不假定方差齐性			8.085	1 064.66	0.000	2.32	0.287	1.757	2.883

输出包括两组不同的检验结果。到底要用哪一组结果,这取决于 Levene 氏齐性方差检验结果的 F 值的显著性。如果 F 值显著,即 $p<0.05$,表明齐性方差假定不成立。这种情况下,就要看与"不假定方差齐性"对应的那组检验结果。如果齐性方差检验结果不显著,即 $p>0.05$,表明齐性方差假定成立。这种情况下,就要看与"假定方差齐性"对应的那组检验结果。从第二个分表中我们可以清楚地看出,Levene 氏齐性方差检验结果的 F 值的显著性远远大于 0.05($p=0.559$),因此,我们可以假定男性样本的首婚年龄分布方差与女性样本的首婚年龄分布方差大体相等。

不过,至此,我们的所有分析还没有回答均值差异的问题。Levene 氏 F 检验只是我们检验过程中的一个插曲,目的仅限于查看一下两个样本的方差差异情况。插曲之后,我们就可以确定均值之间的差异情况。可以注意到,有两个 t 值,一个等于 8.066,与"假定方差齐性"对应,另一个等于 8.085,与"不假定方差齐性"对应。Levene 氏齐性方差检验结果告诉我们,齐性方差假定成立,因此,我们取与"假定方差齐性"对应的那个 t 值。可以看出,与这个 t 值对应的显著性水平是 0.000,或者可简捷地说,$p<0.001$。

得出结论,解释结果:根据统计分析结果我们可以得出这样的结论:对于这样大小的样本以及它们的标准差和方差,单凭偶然性,我们获得 8.066 这么大 t 值的可能性不到千分之一(即 $p<0.001$),于是我们说,男性平均首婚年龄和女性平均首婚年龄之间的差异显著。换句话说,男性结婚平均要比女性晚,或者女性结婚平均要比男性早。但这组数据并没有告诉我们具体某一对夫妻的情况,它所告诉我们的是由大量夫妻组成的团体的情况。

表 4.6 中的数据并没有对两性首婚年龄差异提供解释。文化常例、传统、性别角色、生育率以及其他理论,都可以帮助我们了解人们的首婚年龄问题。

练习框 4.3

自购房者的年龄大,还是租房者的年龄大? 让我们看一看《普通社会调查》结果所揭示的情况。

组统计量

	产权	样本人数	均值	标准差	标准误
年龄	自购	659	49.29	16.991	0.662
	租赁	322	40.28	16.942	0.944

独立样本 t 检验结果

		Levene 氏方差齐性检验		均值差异 t 检验					95%置信区间	
		F	*Sig.*	t	df	*Sig* (双尾)	均值差	标准误	下界	上界
年龄	假定方差齐性	0.438	0.508	7.809	979	0.000	9.01	1.154	6.748	11.278
	不假定方差齐性			7.817	638.8	0.000	9.01	1.153	6.749	11.278

问　题

1. 哪些是变量? 它们的测量水平分别是什么?
2. 所检验的假设是什么?
3. 为什么这里用独立样本 t 检验是合适的?
4. 用文字描述这个表格所展示的内容。例如:Levene 氏方差齐性检验告诉了我们些什么? 我们应该用哪个 t 值?
5. 用日常生活语言对该 SPSS 输出结果作个结论。
6. 你能否得出结论说,自购房者与租赁房者有年龄差异,或者无年龄差异? 还有些什么别的信息可以帮助你作出结论?
7. 这些结果的可能解释是什么? 有没有理论或者以前的研究可以引导你解读该研究中的发现?

方差分析

电视人人都看。但是,每天看电视的时间长短是否取决于某些特殊的因素? 我们一起来看一下表 4.7 中所呈现的有关

表 4.7

每天看电视时间描述统计结果

文化程度	人数	均值	标准差	标准误	95% 置信区间		最小值	最大值
					下界	上界		
高中以下	276	3.93	2.820	0.170	3.59	4.26	0	24
高中	774	2.96	2.137	0.077	2.81	3.11	0	20
大专	90	2.63	2.564	0.270	2.10	3.17	0	20
大学	233	2.06	1.220	0.080	1.90	2.22	0	10
研究生	113	1.83	1.202	0.113	1.61	2.06	0	6
总计	1 486	2.89	2.232	0.058	2.78	3.01	0	24

每天看电视时间方差分析结果

来源	*SS*	*df*	*MS*	*F*	*p*
组间	593.948	4	148.487	32.305	0.000
组内	6 807.252	1 481	4.596		
总计	7 401.201	1 485			

文化程度与每天看电视时间长短之间的关系。

找出变量:看电视的时间长短是一种定距/定比水平的量数,适合计算均值。表示教育水平的五个类别(等级)是定序量数。因为我们要比较五种文化程度被调查者之间每日平均看电视的时间长短差异,方差分析是适当的显著性检验方法。

解读图表:第一个分表呈现的是五组被调查者每天看电视时间长短的描述统计结果。例如,774 位高中文化程度的被调查者每天看电视的平均时间约 3 小时(确切为 2.96 小时)。该组中,至少有一位被调查者每天看 0 小时的电视 (最小值),至少还有一位被调查者每天看 20 小时的电视(最大值)。标准差用来刻画每一组被调查者每天看电视时间的离中分散程度,是组内变异的量度。高中以下文化程度的 276 位被调查者每天看电视时间的全距最大(0 ~ 24),这组的标准差也表明了这一点。一个人每天真的能看 24 小时的电视吗?

从数据中就可以看出,文化程度越低的人,每天看电视的时间越长;文化程度越高的人,每天看电视的时间越短。为了检验这些均值差异是否显著,使用了单向方差分析统计方法。F 值(32.3)达到 0.000 的显著性,即 $p < 0.001$。这一信息表明,五个被调查者组间的变异方差比组内的变异方差大。看上去,每天看电视的时间长短似乎取决于被调查者的文化程度。

注意:组间变异的自由度等于被调查者组数(5)减 1,即 4;组内变异的自由度等于每组被调查者人数减 1 所得差的和,或者 5 组被调查者总数减组数,即 1 481。学术期刊文章中将以这样的方式报告分析结果:$F(4,1481) = 32.3, p < 0.001$。从表中可以看出,均方($MS$)等于平方和($SS$)除以相应的自由度数($df$),如 593.948/4 = 148.487。$F$ 值就是组间均方除以组内均方,即 148.487/4.596 = 32.305。

得出结论,解释结果:根据 F 检验结果的显著性水平,我们可以得出:五组被调查者在看电视时间上的均值差异达到预设的统计显著性水平。文化程度最低的,平均看电视最多;文化程度最高的,平均看电视最少。

这些数据可能是解释看电视时间多少的原因的一个尝试，文化程度是一个很重要的指标。但是，为什么这两个变量之间会有联系呢？难道是因为文化程度高的人花在看书上的时间更长？或者不同文化程度的人所从事的工作会影响他们休闲时间的长短？表4.7中的数据并没有提供有关这些问题的答案，不过，文献回顾和进一步的数据分析可以提供更多的解释。

练习框4.4

下表呈现的是大学四个不同专业学生平均级点分的比较结果。

描述统计结果

CUMGPA03

	人数	均值	标准差	标准误	95%置信区间 下界	95%置信区间 上界	最低	最高
心理学	129	3.2429	0.39660	0.03492	3.1738	3.3120	2.0	3.9
社会学	87	3.2624	0.40105	0.04300	3.1769	3.3479	1.8	3.9
政治学	79	3.1377	0.37103	0.04174	3.0546	3.2208	2.1	3.9
经济学	43	3.0540	0.36729	0.05601	2.9410	3.1671	2.0	3.8
总计	338	3.1993	0.39345	0.02140	3.1572	3.2414	1.8	3.9

方差分析结果

CUMGPA03

	SS	*df*	*MS*	*F*	*Sig.*
组间	1.799	3	0.600	3.976	0.008
组内	50.369	334	0.151		
总计	52.168	337			

问　题

1. 哪些是变量？它们的测量水平分别是什么？
2. 所检验的假设是什么？
3. 为什么这里用方差分析是合适的？

4. 用文字描述这个表格所展示的内容。
5. 用日常生活语言对该表格所呈现的结果作个结论。
6. 你能否得出结论说,不同专业学生的平均级点分之间有差异,或者无差异? 还有些什么别的信息可以帮助你作出结论?
7. 这些结果的可能解释是什么? 有没有理论或者以前的研究可以引导你解读该研究中的发现?

t 检验与方差分析小结

很多研究问题都是关于跨时比较或跨组比较的。其中一种分析方法就是对不同时间或不同组的均值进行比较。能做这样比较的变量,都必须是在定距/定比水平测量的,或者在适合计算算术平均数的定序水平测量的,像看上去定距的利克特氏量标上的量。有一点非常重要,应该时刻牢记,那就是:均值本身只是一个总结性统计量,仅用均值,有关分布的其他信息就会丢失,标准差提供了关于分布的更多描述信息。前面我们已经看到,检验均值差异时,已经把各组数据分布的不同标准差考虑了进去。

如果均值比较是在两个互斥的组间进行,要采用独立样本 t 检验;如果均值比较是在同一个组的两个不同测量结果间进行,采用配对样本 t 检验才是合适的,因为这种方法把同组被试在两个不同测量结果间的相关考虑了进去。

如果均值比较是在三个或三个以上的组间进行,那就要用单向方差分析技术,因为单向方差分析提供的 F 统计量可以帮助我们确定组间的变异方差是不是大于组内的变异方差。均值差异的显著性检验是一种常见的统计技术,是皮尔逊相关 r 和交叉表所提供的信息的一种补充。

第5章 解读回归数据

对于研究问题的回答，通常包括对两个以上变量之间关系的分析。有些时候，研究所关注的是如何同时通过几个变量来解释、预测或描述另一个变量。多元回归分析，又名常最小二乘[①]线性回归，便是一种估量两个及两个以上自变量之间关系以及这些自变量与因变量之间相关关系的常用统计技术。

① “最小二乘法”分“常最小二乘法”和“偏最小二乘法”。在现行的汉语统计学文献中，一般用“最小二乘”指“常最小二乘法”。为了突出两种最小二乘法之间的差异，译文中用了“常最小二乘法”。

定 义

■ 复相关以皮尔逊相关为基础，R 是复相关系数，取值在 0 至 1 之间。该系数是因变量与两个及两个以上自变量关联性强度的量度。确定系数 R^2 表示因变量方差能被几个自变量方差共同解释的比例。有时，也把自变量叫做预测变量或准则变量。

■ **回归系数**：B 与 *Beta*（即 β）表示由每一个自变量引起的因变量的变化率。如果把回归关系式用一条回归线图像表示的话，B 实际上就是这条回归线的斜率。B 是非标准化回归系数，β 是标准化回归系数，因此，只有当各变量的测量结果使用同样的单位时，不同自变量的回归系数才具有可比性，否则，使用标准化回归系数 β 更合适。回归系数表明了相应的自变量在解释因变量变异方差时相对于其他自变量的正负方向和权重。

假 定

■ 回归分析假定，因变量和自变量之间存在线性关系。

■ 当各自变量呈现低度相关或低多元共线性时，即当自变量之间不是强相关时，采用回归分析的效果最佳。

■ 变量为定距/定比水平，或看上去是定距水平，或哑变量（典型取值为 0 和 1 的二分变量）。

取自学术论文的真实数据

尽管许多市场营销公司、民意调查机构和其他研究中心都

经常使用回归分析方法,但是,回归分析结果的有关统计量很少见诸于大众报刊。不过,在学术刊物和其他研究性出版物中,回归分析及其有关统计量则是经常出现的。

探索多自变量和单因变量之间关系的回归方式有多种,其中最普遍的两种方式是一次性全部进入法和逐步法。全部进入法按照研究者对自变量的排序对每个自变量一一进行检验。不论显著与否,每个自变量都要进入最后的回归模型。研究者必须对最后模型加以解释,并决定哪些变量具有统计显著性。另一个选择就是使用逐步回归法。这种方法可能一步完成,也可能多步完成,这取决于到底有多少个自变量与因变量显著相关。

当所有变量按照其重要性都进入模型,或者,不再有显著相关的自变量可以进入模型时,回归进程终止。所有这些工作都可以由计算机程序来完成,即由统计软件来决定回归分析中哪些变量是重要的。

一次性全部进入回归法

表5.1呈现的是一项关于选举中为什么某特定州的县更倾向于支持环境保护措施的研究结果。研究的数据并不是基于个体的投票数据,而是基于县的总体投票数据,因此,分析的单位不是个体"人",而是个体"县"。尽管原研究分析了几个州的数据,但在此我们重点观察密歇根州的情况。请注意,所有的变量都进入了分析,其中也包括那些不具有统计显著性的变量。

找出变量:因变量是赞成环保措施者占投票总人数的百分比,抽样单位是县。所以,表5.1是密歇根州83个县的投票情况。投票所涉及的措施包括"公园"(捐款建设更多的州公园)、"狩猎"(禁止在春天猎杀黑熊)、"野生动物"(控制和规范野生动物的呈交和领养)、"债券"(批准设立专项债券以净化受污染场所,提高水的质量)。

表 5.1 密歇根州 1990—2000 年投票支持各项措施的回归分析结果
（括号中的是标准化回归系数）

	公园 1994 NE	狩猎 1996 NE	野生动物 1996 NE	债券 1998 NE
人口密度	0.000	0.012**	-0.005	0.012*
	(0.00)	(0.28)	(-0.21)	(0.22)
城市/50 000 +	0.88	3.36	-0.59	0.31
	(0.05)	(0.18)	(-0.05)	(0.01)
在资源型产业中就业的比例	-0.007	-0.07	-0.06	-0.07
	(-0.001)	(-0.08)	(-0.11)	(-0.07)
收入中位数/千美元	0.04**	0.007	0.02*	0.49**
	(0.40)	(0.07)	(0.36)	(0.43)
教育程度	0.22	0.56**	-0.39**	0.35**
	(0.19)	(0.50)	(-0.56)	(0.27)
投共和党票者比例	-0.10	-0.20*	-0.14*	-0.12
	(-0.10)	(-0.19)	(-0.23)	(-0.10)
35 岁以下者比例	0.25	-0.17	0.12	0.17
	(0.18)	(-0.13)	(0.14)	(0.11)
调整后确定系数 Adj-R^2	0.496	0.566	0.231	0.643
数量	83	83	83	83

注：* $p<0.05$，** $p<0.01$。

来源：Salka (2003)　[研究检索索引号:9750040]

自变量包括各县的人口密度（即每平方公里的人口数量，人口密度越小，表明该县农村化程度越高），即该县是否有 5 万人及以上人口的城市使用了哑变量，居民中在资源相关产业（如种植/养殖、木材业、采矿业等）中从业的比例，县收入的中位数，25 岁以上居民中大学文化程度者所占的比例，1996 年总统大选（罗伯特 · 道尔）中投共和党票者的比例，居民中年龄在 35 岁以下者的比例。所有这些变量的测量都在定距/定比水平。

在这个例子中,由于因变量和所有自变量的测量都是定距/定比水平,且研究目的是解释各县在赞成还是反对治理措施投票上的变异情况,因此,多元回归是一种恰当的统计分析方法。

解读图表:对于回归分析,回归系数(包括非标准化回归系数 B 和标准化回归系数 β)和确定系数 R^2,在绝大多数情况下已经提供了我们所需要的信息。虽然统计软件的输出结果大都包括对回归方程的显著性(F)检验以及每个回归系数的显著性(t)检验,但是,在出版发表研究结果时,有时并不报告这些统计量。回归系数告诉我们每个自变量与因变量相关系数的权重(程度)和方向,确定系数 R^2 表明方程中所有自变量一起所能解释因变量变异方差的比例。

表 5.1 既呈现了非标准化回归系数,也呈现了对应的标准化回归系数(放在括号中)。标准化回归系数可以用来比较自变量在四个回归(即关于四项措施的投票结果)中的相对重要性。切记,对于相同单位的自变量(如“人口密度”),可以比较它们的非标准化回归系数 B,对于不同单位的自变量(如“收入中位数”和“35 岁以下居民的百分比”),则最好比较它们的标准化回归系数 β。

除了“公园”项外,对于其他项措施的回归,分别都有三个自变量的回归系数达到设定的统计显著性水平。例如,“人口密度”、“收入”以及“教育程度”都与支持设立专项“债券”以净化受污染场所措施显著相关。如果根据标准化回归系数来比较这三个与“债券”显著相关的自变量,其重要性顺序是:“收入”(0.43)、“教育程度”(0.27)、“人口密度”(0.22)。由于这些系数都是正值,在这三个自变量上的取值越高,对设立专项“债券”的支持程度也就越高。

所有七个自变量一起,能解释对这项措施支持变异方差的 0.643(R^2),即 64.3%。统计软件既计算确定系数 R^2,也计算调整后的确定系数 Adj-R^2。当自变量为多个时,推荐用调整后的确定系数。

"收入"和"教育程度" 看上去对三项措施的投票结果都具有较强的预测力。注意,表中"投共和党票者比例"的回归系数都是负值,表明在这些县中,支持共和党的票数越多,支持环境保护措施的票数就越少。除对"野生动物"一项措施外,七个自变量加起来能够解释的其他环保措施的变异方差的比例约有一半或者更多。例如,支持"狩猎"措施的变异方差的 56.6% 可由各县的"教育程度"($\beta = 0.50$)、"人口密度"(0.28)以及"投共和党票者比例"(−0.19)解释。

得出结论,解释结果:"教育程度"和"收入"这两个自变量看上去对三项措施的投票结果都具有最强的预测力;"政党"和"人口密度"与两项措施的投票结果显著相关。总的来说,在那些教育程度越高、收入越高、城市化程度越高、支持共和党的人数越少的县中,改善环境措施得到支持的可能性就越大。例如,"狩猎"措施的支持者主要来自那些教育程度高、城市化程度高以及共和党支持者少的诸县。

对于解释和预测结果,回归分析是一种很有用的统计技术。本研究就是通过回归分析技术了解人们为什么支持或反对环境保护措施的。教育程度、收入水平、城市化程度(城镇密度)、政党支持取向为什么与环境保护措施相关呢? 这还需要进一步的解释。表 5.1 中提供的那些数据还不能回答这个问题。和我们所碰到的其他研究一样,每一种数据分析方法只能回答一些问题,但在回答这些问题的同时,也会提出其他的问题待进一步的研究来回答。

逐步回归法

另一种多元线性回归分析方法是逐步回归法。在运用这种回归方法时,计算机程序按照自变量的重要性顺序选择达到统计显著性水平的自变量。回归进程的第一步是找出与因变量相关最高的自变量;第二步是找出与因变量相关第二高的自变量(如果有的话),并根据这两个变量计算回归结果数据;重复上述过程直到所有达到统计显著性水平的自变量都进入回

归方程;最后呈现出最佳回归模型。

偶尔,研究者也会选择从另一个方向开始,那就是,第一步让程序把所有变量都包括进去,然后每次去除相关最弱的变量,直到得到最后的回归模型。SPSS 统计软件包就有这种“后退式”逐步回归分析功能。表5.2 呈现的是一项“后退式”逐步回归分析的结果,所涉及的研究问题是,根据对 86 名师范生的调查结果探索如何消除计算机使用焦虑的方法。

表 5.2 用全部变量预测计算机使用焦虑后测水平的逐步回归结果汇总

变　量	非标准系数 B	B 的标准误	标准系数 β	t 值
第一步				
计算机自信心	-0.93	0.23	-0.45	-3.96*
计算机经验	-0.07	0.17	-0.03	-0.39
计算机知识	-0.44	0.19	-0.19	-2.31**
计算机喜爱程度	-0.53	0.23	-0.24	-2.34**
计算机有用程度	0.29	0.29	0.08	1.01
控制点	-0.14	0.19	-0.05	- 0.73
潜质焦虑	0.33	0.08	0.28	3.93*
第四步				
计算机自信心	-0.87	0.23	-0.42	-3.84*
计算机知识	-0.51	0.16	-0.22	-3.19***
计算机喜爱程度	-0.45	0.22	-0.21	-2.10**
潜质焦虑	0.33	0.08	0.28	4.22*

注:对于第一步到第三步,$R=0.84$,$R^2=0.70$;对于第四步,$R=0.83$,$R^2=0.69$,N=86。

来源:Rovai and Childress (2002/2003)　[研究检索索引号:8948093]

找出变量:这项研究的因变量是学生对使用计算机的焦虑程度,数据是在计算机入门课结束时测得的,测量工具是一个由 24 个项目构成的定距/定比焦虑量表,最低分 20 分,最高分 80 分。7 个自变量也是定距/定比水平的数据,分数越高,表示他们对使用计算机越自信,越喜欢使用计算机,越认为计算机对工作有用,使用计算机的经验越丰富且能力越强(从每周使用计算机的小时数、使用计算机的年数来看),关于计算机方面的知识(从一无所知到懂得计算机网络、编程和软件方面的实

用知识）越多，个性特质焦虑程度以及根据标准化内—外控制点量计①（“人们生活中的不愉快事情或多或少都是因为运气不佳”）所测得的外控制程度就越高。所有变量的测量结果都在定距/定比水平，因此，适合作线性回归分析。

解读图表：在该实例中，研究者选择了使用逐步回归分析，而且是后退式去除法，每次去除一个解释力最弱的变量。所以，采用的回归分析步骤是先将全部自变量都引入模型，到第四步结束时，模型中仅剩下几个对因变量有显著预测力的自变量。

发表结果时，研究者通常只给出第一步和最后一步的分析结果。从第一步的分析结果中可以看出，有四个自变量达到预设的统计显著性水平。到第四步结束时，模型中剩下的也正好是这四个自变量。由于和其他自变量间的关系，有些自变量的预测力从第一步的分析结果看并不显著，但随着其他自变量的逐个去除，这些自变量对因变量的预测力可能会变得显著。

表 5.2 呈现了非标准化回归系数（B）及其标准误（意思和标准差类似）、标准化回归系数（*Beta* 或 β），以及检验各回归系数显著性的相应 t 值。数值右上角的星号表示单凭偶然性获得该大小 t 值的显著性水平，仅标注了与各显著性自变量相对应的 t 值的显著性水平。由于各自变量测量结果的取值范围不一（例如，师范生在计算机使用经验上的分数，取值范围是 0～25，特质焦虑测量结果的取值范围是 20～80），所以，要使用标准化 β 系数来比较每个变量对于预测计算机焦虑的重要性。

表 5.2 的脚注告诉我们，第一步回归结果的复相关系数为 0.84，确定系数 R^2 为 0.70。也就是说，对于这个有 86 个学生的样本，7 个自变量（包括未达到统计显著性水平的变量）合起来一共解释了学生计算机焦虑变异方差中的 70%。

到了第四步（第 2 步和第 3 步的分析结果没有呈现）结束

① 在有些中文文献中，把“内—外控制点”也称作“内—外心理控制源”。“标准化内—外控制点量计”是一个测定控制点内—外程度的标准化心理测量工具。

时,模型中仅包括对因变量有显著预测力的四个自变量,这四个自变量一起与因变量的复相关系数为0.83,解释了因变量变异方差的69%。大家要注意,非标准化回归系数(B)和标准化回归系数(β)在各步分析结果中是不同的。如果把四个步骤的分析结果都呈现出来的话,读者就可以看到,由于各变量间相互关系的不同,每步分析的结果都有变化。对于回归分析,最为理想的情况是,自变量之间呈低度相关(低多元共线性)。在分析进程中,随着变量的剔除或加入(对于"前进式"加入法而言),其他变量所承担的解释和预测因变量方差的工作份额必然会有所变化。

大家同时也应该注意到,从第一步的结果到第四步的结果,R^2 几乎是一样的,这表明三个对因变量解释力弱的变量对回归总结果并无多大贡献,因此,删除它们并没有削弱总体复相关系数(R)和对因变量变异方差解释的比例(R^2)。

得出结论,解释结果:复相关系数0.83已达到强相关程度,表明自变量一起总共解释了计算机焦虑变异方差的绝大部分。特别是,与其他专业的学生相比,师范生由于对自身的计算机技能缺乏信心,对计算机方面的知识了解较少,对计算机喜欢程度较低,加之属于更焦急型的人格,因此,在他们刚开始教学生涯时,更倾向于表现出较高的焦虑水平。注意,在保留下来的四个自变量中,三个的系数都是负值,说明这三个自变量与因变量之间的关系是相反的。

特质焦虑假定,个体存在一种人格性征,这种性征与计算机焦虑相关,这一点符合我们的直觉。同样,那些对自己的计算机技能信心不高、计算机知识缺乏且又不喜欢使用计算机的学生,在成为教师,初次在课堂上使用计算机时有更多的焦虑感,这种情况也是可以理解的。很清楚,师范教育的任务就是帮助那些未来的教师建立他们在未来职业生涯中使用计算机的信心,并使他们获得更多的关于计算机的知识和技能。

练习框 5.1

对一个由996 人构成的被调查样本进行了一次关于性行为和婚姻功能方面的电话调查。婚姻功能分数由利克特氏量计测得，量计共有 10 个问题，总分是各问题得分之和。问题涉及婚姻关系运行状况、幸福感程度、吵嘴的次数、亲吻配偶的频度、是否后悔结婚等。表中呈现的 9 个预测变量关注的是性行为及其满意程度。最后一个问题是一个 9 等级分的题目，高分表示高满意程度。其余的项目都属于真/伪选择型题目，都是测量关于性行为的观点的，高分表示观点为“真”。

男女性行为满意度多元回归分析结果

预测变量	60 岁以下		60 岁及以上	
	女性 (n = 435)	男性 (n = 368)	女性 (n = 64)	男性 (n = 116)
性生活持续时间不够长	-0.03	0.08	0.19	-0.09
性生活次数不够多	0.03	0.00	0.11	-0.14
对性伴侣满意	0.15**	0.05	0.28*	0.16
性交流通畅	0.18**	0.17**	0.23	0.15
多样性不够	-0.08	-0.17**	0.00	0.01
互相爱抚	0.09*	0.09*	0.04	0.06
令人满足的性高潮	0.00	0.08	0.16	0.04
专注于性动作	0.03	0.00	0.12	0.03
对性关系满意	0.35**	0.28**	0.06	0.20**
R^2	0.38**	0.28**	0.34*	0.25**

注：* $p<0.05$，** $p<0.01$。

来源：Trudel (2002) [研究检索索引号：6810589]

问　题

1. 自变量和因变量分别是哪些？它们的测量水平(定类、定序、定距或定比)分别是什么？
2. 所检验的假设是什么？
3. 为什么这里用多元回归统计技术是合适的？
4. 用文字描述这个表格所展示的内容。例如:哪些自变量显著地预测了60岁以下男子的婚姻满意度？哪些自变量又显著地预测了60岁以上女子的婚姻满意度？等等。四个不同的回归模型分别解释了因变量变异方差的百分之多少？在这四个回归模型中,负值回归系数的意思是什么？那个“多样性不够”自变量的负值回归系数又是什么意思？
5. 用日常生活语言对该结果作个结论。
6. 你能否得出结论说,这里有年龄差异,或者无年龄差异？还有没有什么别的信息可以帮助你作出结论？
7. 这些结果的可能解释是什么？有没有理论或者以前的研究可以引导你解读该研究中的发现？

SPSS 输出

进入式回归法

在美国,并不是每个人都上大学,大学毕业,然后继续研究生学习。事实上,大部分美国人都没有大学文凭。那么,哪些因素可以预测人们接受更多教育的可能性呢？

表5.3是来自美国《普通社会调查》的一项多元回归分析的结果,目的是了解并用进入式回归法分析所保留的自变量来预测普通美国人接受教育的年限。

表 5.3

回归模型

模型	R	R^2	Adj-R^2	估计量的标准差
1	0.612[a]	0.374	0.370	2.497

注：a 预测变量：(常数)、居住地规模、被调查者的性别及兄弟姐妹数、被调查者的社会经济指标。

方差分析[b]

模型		平方和	自由度	均方	F 值	显著性
1	回归	2 617.305	4	654.326	104.976	0.000[a]
	残差	4 381.852	703	6.233		
	总计	6 999.157	707			

注：a. 预测变量：(常数)、居住地规模、被调查者的性别及兄弟姐妹数、被调查者的社会经济指标。

b. 因变量：受教育的最高年限。

回归系数[a]

模型		非标准系数		标准系数		
		B	标准误	β	t 值	显著性
1	(常量)	8.511	0.455		18.704	0.000
	性别	0.384	0.191	0.061	2.015	0.044
	社会经济指标	0.094	0.005	0.562	18.057	0.000
	兄弟姐妹数	-0.132	0.031	-0.131	-4.255	0.000
	居住地规模	0.000	0.000	0.059	1.952	0.051

注：a. 因变量：受教育的最高年限。

找出变量："受教育的最高年限"是定距/定比量数，通过询问调查对象所完成学校教育的年限而获得。由于研究的目的是了解什么因素能够预测或者解释这 707 名调查对象受教育最高年限上的变异，所以，受教育年限是因变量。预测变量(即自变量)包括性别(二分定类变量，1 = 男，2 = 女)、社会经济指

标(定距/定比变量,基于教育程度、收入及职业声望,指标越高,社会经济地位就越高)、兄弟姐妹人数(定距/定比变量)、居住地规模(定距/定比变量,数据压缩了 1 000 倍)。定距/定比数据和二分数据都适合进行多元回归分析。

解读图表:表 5.3 中的第一个分表呈现了回归模型的关键组成部分,即所有自变量对于因变量的复相关系数 R。为了确定几个自变量一起所解释的因变量变异方差的比例,计算了确定系数 R^2。当自变量数目很多时,应该用调整(后的)确定系数 Adj-R^2,因为"调整确定系数"考虑了每个自变量与进入模型中的其他自变量之间的关系影响。确定系数为 0.374 的意思是,性别、社会经济指标、兄弟姐妹数、居住地规模这些因素联合起来,一共解释了调查对象受教育年限变异方差的 37.4%。

表中最后一列中的值是估计量的标准差。该统计量与标准差解释类似。它是实际分数偏离回归预测分数平均程度的量度,标准差的值越小,预测的误差就越小。

第二个分表报道的是关于回归方程显著性的检验结果。从检验结果中我们可以得知,回归线是否与数据拟合,或者在自变量与因变量之间是否存在某种线性关系。表中还报道了 F 值及其相应的显著性水平。如果 $p<0.05$(当然也可以根据需要设定为0.01 甚至0.001),我们就可以得出结论说:回归达到统计显著性水平。

最后,第三个分表报道了每个自变量对于预测因变量的实际效果,这些值也许是最为重要的信息。对于测量单位相同的变量,非标准系数(B)具有可比性;对于测量单位不同的变量,标准系数(β)才具有可比性。由于本研究中各变量的测量单位不同,所以应该比较标准系数。与皮尔逊相关系数类似,系数的负方向用"-"表示,相关联程度由系数值的大小表示。系数的具体意思就是回归线关于每个自变量的斜率。为了检验各回归系数不是0,计算并报道了与每个回归系数对应的 t 值及其显著性水平。在逐步回归模型中,未达到0.05 显著水平

的预测变量，在最终模型中都予以删除。不过，本研究用的是全部进入式进程，所以，不管回归系数是否达到设定的统计显著性水平，统统予以报道。如果使用逐步回归法，报道时只报道那些达到预设显著性水平的回归系数。

报道中的第一个预测变量是“常数”，即回归线与 Y 轴的交点。当 $x=0$ 时，也就是说，当我们对自变量一无所知时，“常数”是对因变量的最优预测值。呈现结果时通常不呈现“常数”，只是在描绘回归线，或者使用回归方程预测分数时才报告“常数”的值。

变量性别、社会经济地位、兄弟姐妹数的回归系数都达到了统计显著性（性别 $p<0.05$，其他变量 $p<0.001$）要求，而居住地规模的显著性是 0.051，几乎达到一般设定的统计显著性水平（即 $p<0.05$）。我们可以看出，“社会经济指标”的标准回归系数最大，$\beta=0.562$，其次是“兄弟姐妹数”（-0.131）和性别（0.061），最后是“居住地规模”（0.059）。这些自变量一起，解释了调查对象受教育年限变异方差的 37.4%。

得出结论，解释结果：用文字描述，我们可以这样说：那些社会经济地位高、兄弟姐妹数少，且来自大地方的女性，倾向于接受更长时间的教育。除了“兄弟姐妹数”自变量外，其他所有自变量的系数都为正值。这就意味着，在社会经济地位（职业声望和收入）、性别（1 = 男，2 = 女）和居住地规模变量上的值越大，且在兄弟姐妹数变量上的值越小，所受教育的年限则可能越长。不过，需要注意的是，尽管有 37.4% 的变异方差得到了解释，但还有 62.6% 的变异方差仍未得到解释。很显然，还有其他因素能够预测受教育的年限。

我们的确知道，女性的大学毕业率高于男性，从事高声誉工作和高收入工作的人倾向于接受更多的教育，教育程度高的人倾向于居住在更大的城区。为什么兄弟姐妹数更少者受教育的年限更长呢？原因可能是，兄弟姐妹少，父母支付孩子接受更长时间教育的费用的能力就更大。你如何看这个问题呢？你可以在美国《普通社会调查》的其他问题的回答中找到这个

问题的解释,你也可以把它作为一个题目去研究,撰写学位论文或撰写论文发表。

练习框 5.2

下表中呈现的是一张 SPSS 回归分析的结果输出,方法是全部进入式,被试是 762 名大学生,分析的目的是找出大学平均级点分的最佳预测变量。不用说,大学生的在校学业表现受很多因素的影响。不过,本研究只考虑了以下几个变量的影响:高中平均级点分、SAT 总分、高中的名次、性别(1 = 女,2 = 男)。

回归模型

模型	R	R^2	Adj-R^2	估计量的标准差
1	0.413[a]	0.170	0.166	0.622 20

注:a. 预测变量:(常数)、性别、SAT 总分、高中名次、高中平均级点分。

方差分析[b]

模型		平方和	自由度	均方	F 值	显著性
1	回归	60.184	4	15.046	38.865	0.000[a]
	残差	293.446	758	0.387		
	总计	353.630	762			

注:a. 预测变量:(常数)、性别、SAT 总分、高中名次、高中平均级点分。
b. 因变量:大学平均级点分。

回归系数[a]

模型		非标准化系数		标准化系数	t 值	显著性
		B	标准误	β		
1	(常量)	1.137	0.333		3.414	0.001
	高中平均级点分	0.378	0.074	0.262	5.123	0.000
	高中名次	-0.016	0.019	-0.042	-0.841	0.401
	SAT 总分	0.001	0.000	0.157	4.546	0.000
	性别	-0.226	0.048	-0.162	-4.727	0.000

注:a. 因变量:大学平均级点分。

问 题

1. 自变量和因变量分别是哪些？它们的测量水平分别是什么？
2. 所检验的假设是什么？
3. 为什么这里用多元回归统计技术是合适的？
4. 用文字描述这个表格所展示的内容。例如：R 是什么意思？R^2 又是什么意思？F 检验结果告诉了我们关于回归模型的什么信息？哪些自变量是因变量的良好预测变量？B 系数和 β 系数之间的区别是什么？为什么有的回归系数是负的？
5. 用日常生活语言对该结果作个结论。
6. 你能否得出结论说，自变量和因变量之间存在某种关系？还有些什么别的信息可以帮助你作出结论？
7. 这些结果的可能解释是什么？有没有理论或者以前的研究可以引导你解读该研究中的发现？

逐步回归法

有些时候，我们也许并不知道哪些自变量可能与因变量有关系，也没有任何关于自变量和因变量之间关系的理论。这种情况下，多元逐步回归法可以帮助探明哪些自变量与因变量关联最为密切，以及自变量之间是如何交互影响的。

我们现在来看表 5.4 中的内容。该项研究试图探究为什么人们生育孩子的数量存在差异。

找出变量：生育孩子的数量是一个离散型定距/定比变量，尽管最后一个值代表生育了 8 个及以上孩子。考虑到生育 8 个及以上孩子的人数实在是太少，这个变量基本上还是定距/定比水平的，所以适合作回归分析。同样，调查对象的年龄、他们首婚时的年龄以及兄弟姐妹人数这些自变量都是定距/定比水平的变量，所以，可以进行回归分析。

解读图表：表 5.4 由四个分表构成，简要报告了逐步回归分析的过程和结果。第一个分表呈现了每步回归的复相关系数 R 和确定系数 R^2。请注意，本例采取三步回归，每步对应一个回归模型。第一步回归的 R 值为 0.301，到最后一步时，R 值

表 5.4

回归模型

模型	R	R^2	Adj-R^2	估计量的标准差
1	0.301[a]	0.091	0.089	1.511
2	0.412[b]	0.170	0.167	1.446
3	0.438[c]	0.191	0.187	1.428

注:a. 预测变量:(常数)、调查对象的年龄。
b. 预测变量:(常数)、调查对象的年龄、首婚时年龄。
c. 预测变量:(常数)、调查对象的年龄、首婚时年龄、兄弟姐妹人数。

方差分析[d]

模型		平方和	自由度	均方	F 值	显著性
1	回归	138.801	1	138.801	60.767	0.000[a]
	残差	1 388.755	608	2.284		
	总计	1 527.556	609			
2	回归	259.206	2	129.603	62.025	0.000[b]
	残差	1 268.350	607	2.090		
	总计	1 527.556	609			
3	回归	292.450	3	97.483	47.830	0.000[c]
	残差	1 235.105	606	2.038		
	总计	1 527.556	609			

注:a. 预测变量:(常数)、调查对象的年龄。
b. 预测变量:(常数)、调查对象的年龄、首婚时年龄。
c. 预测变量:(常数)、调查对象的年龄、首婚时年龄、兄弟姐妹人数。
d. 因变量:生育孩子的数量。

续表

回归系数

模型		非标准系数		标准系数		
		B	标准差	β	t 值	显著性
1	(常数)	0.805	0.197		4.095	0.000
	年龄	0.029	0.004	0.301	7.795	0.000
2	(常数)	2.719	0.315		8.645	0.000
	年龄	0.031	0.004	0.332	8.931	0.000
	首婚时年龄	-0.090	0.012	-0.282	-7.591	0.000
3	(常数)	2.491	0.316		7.887	0.000
	年龄	0.030	0.003	0.315	8.530	0.000
	首婚时年龄	-0.089	0.012	-0.279	-7.598	0.000
	兄弟姐妹数	0.072	0.018	0.148	4.039	0.000

注:因变量:生育孩子的数量。

删除变量[d]

模型					偏相关	共线性
		β	t	显著性	系数	容忍度
1	兄弟姐妹数	0.154[a]	4.018	0.000	0.161	0.988
	受教育年限	-0.136[a]	-3.381	0.001	-0.136	0.904
	首婚时年龄	-0.282[a]	-7.591	0.000	-0.294	0.988
	居住地规模	-0.046[a]	-1.190	0.234	-0.048	1.000
2	兄弟姐妹数	0.148[b]	4.039	0.000	0.162	0.987
	受教育年限	-0.067[b]	-1.675	0.094	-0.068	0.849
	居住地规模	-0.039[b]	-1.065	0.287	-0.043	0.999
3	受教育年限	-0.027[c]	-0.665	0.507	-0.027	0.791
	居住地规模	-0.034[c]	-0.923	0.356	-0.037	0.998

注:a. 预测变量:(常数)、调查对象的年龄。

b. 预测变量:(常数)、调查对象的年龄、首婚时年龄。

c. 预测变量:(常数)、调查对象的年龄、首婚时年龄、兄弟姐妹数。

d. 因变量:生育孩子的数量。

达到0.438,相应的R^2也从大约9%上升为19%。很显然,第三步呈现了一幅更加完善的自变量与因变量之间关系的图画。增加变量可以提高解释因变量变异方差的比例。

第二个分表呈现的是检验每步回归模型显著性的方差分析结果。F值的显著性水平为0.000,表明每步产生的回归方程在$p<0.001$水平上具有统计显著性。通常,论文中并不报告这些信息。该信息是为了引导研究者核验回归模型的显著性。

第三个分表呈现的是各自变量的标准回归系数(β)和非标准回归系数(B)。同时也呈现了相应的t检验值,用于分析相应系数不是0的程度。每步分析都有一个回归模型,总共三个模型。最重要的自然是最后一个模型即第四个分表,因为它包括了所有达到显著性水平的自变量。可以比较一下第三个分表呈现的信息和第四个分表(删除变量)所呈现的信息。后者呈现的是每步分析时未进入模型的自变量的回归系数。

例如,在第一步,"调查对象的年龄"作为预测生育孩子数量的最强变量进入回归方程,其β系数(由于孩子的数量与首婚时年龄的单位不同而缺乏可比性,因此,考虑标准回归系数)为0.301。比较系数0.301和删除变量分表中第一步的回归系数可以看出,0.301是最高值。从第四个分表中还可以发现,在第一步所删除的变量中,变量"首婚时年龄"的t值最大,统计显著性最高,因此,它在第二步分析中进入回归方程。

在第二步分析中,"首婚时年龄"以系数-0.282进入回归方程(负值并不意味相关程度低,它代表的是关联的方向,而不是关联程度)。同时也要注意,在自变量"首婚时年龄"进入回归方程后,"调查对象的年龄"这一变量的β系数发生了变化(从0.301变为0.332)。切记,对于多元回归,几个自变量一起同时解释因变量的变异。所以,"调查对象的年龄"单独与因变量的相关系数是0.301,但是在"首婚时年龄"加入时,它与因变量的相关系数变为0.332。

现在我们再回到第一个分表看一看。我们发现,"调查对象的年龄"和"首婚时年龄"两个自变量一起,与因变量的复相

关系数为 0.412。可见,两个自变量共同作用,比它们任何一个单独作用时都能更好地解释因变量。

最后,在第三步,“兄弟姐妹数”变量进入回归方程。请注意,该变量在被删除变量的第二步分析结果中的系数最高,而且是唯一达到统计显著性水平的变量。所以,在第三步分析中就引入了该变量。第四个分表中还提供了我们通常称作偏相关分析(在保持其他变量不变的条件下,一个变量与因变量之间的相关,即该变量与因变量的净相关)的信息,以及其他适合更高级统计分析的信息。在第三步中,我们可以看到哪些变量依然被排除在回归方程之外。这些被排除在外的变量的 β 系数都没有达到统计显著性水平,所以不能进入最后的回归方程。换句话说,这些被排除在外的变量对于解释或预测因变量的作用甚微。

回归系数(第三个)分表的最后一步(第三步)中的信息和回归模型(第一个)分表中的总结性信息是最重要的信息,因为这些信息是关于哪些自变量对于回归分析有贡献以及贡献程度有多大。通过比较不同自变量的 β 系数(或者在测量单位相同时使用非标准回归系数 B)我们可以看出,按照相关强度顺序,“调查对象的年龄”、“首婚时年龄”、“兄弟姐妹数”与“生育孩子的数量”这个因变量相关。参看第一个分表中的第三步分析结果我们可以看出,三个自变量与因变量的共同相关系数为 0.438,解释了因变量总变异方差的 19.1%。

简而言之,SPSS 逐步回归分析输出的关键信息可以在回归模型(第一个分表)部分和回归系数(第三个分表)部分找到。那里报告了回归系数以及它们与因变量的复相关系数。大部分研究者都报告每个自变量的 β 系数、R 值和 R^2 值。

得出结论,解释结果:表 5.4 中所呈现的结果告诉我们,调查对象目前的年龄、首婚时的年龄以及兄弟姐妹人数对他们为什么生育了不同数量的孩子可以提供解释。具体而言,与那些生育较少孩子的人相比,那些年龄更大、首次结婚时年龄更小且兄弟姐妹人数更多的人,更倾向于生育更多的孩子。“首婚

时年龄”变量前的负号表明，该自变量与因变量呈反向关系，即生育孩子少对应首婚时的年龄更大（结婚迟），生育孩子多对应首婚时的年龄更小（结婚早）。

这些发现很合乎情理，因为那些结婚早而且现在年龄大的调查对象有更长的时间生育孩子。此外，年龄大的一代人倾向于大家庭，调查结果正好反映出了这一倾向。再加入兄弟姐妹人数这一与生育孩子数量有点呈正相关的自变量，我们就可以用这三个变量一起来预测人们生育孩子的数量。然而，这三个变量只解释了人们生育孩子的数量变异方差的 19%，还有 81% 的变异方差未得到解释。很显然，还有许多其他因素可以解释为什么人们生育不同数量的孩子。在相关理论和以往研究的指导下，我们可以通过另外一项回归分析研究找出一些其他的变量。教育程度和居住地规模（作为是在城市或乡村生活的指标）与人们生育孩子的数量关系不大，因此，有必要研究其他因素，例如收入水平、职业选择、种族/民族、宗教等。

练习框 5.3

下面是用 SPSS 软件对《普通社会调查》数据做的一个逐步回归分析的部分结果输出，目的是了解自由党和保守党在政见上的差异。政见的分值从“1 = 极端自由”到“7 = 极端保守”；婚姻状况的分值是：1 = 已婚，2 = 未婚；性别的分值是：1 = 男，2 = 女；居住地规模用所在地的人口数量量度。

模型总结

模型	R	R^2	adj-R^2	估计值的标准差
1	0.127[a]	0.016	0.015	1.356
2	0.168[b]	0.028	0.026	1.349
3	0.195[c]	0.038	0.034	1.343

注：a. 预测变量：（常数）、居住地规模（以千人为单位）。

b. 预测变量：（常数）、居住地规模（以千人为单位）、调查对象的年龄。

c. 预测变量：（常数）、居住地规模（以千人为单位）、调查对象的年龄、婚否。

方差分析[d]

模型		平方和	自由度	均方	F 值	显著性
1	回归	21.823	1	21.823	11.863	0.001[a]
	残差	1 320.822	718	1.840		
	总计	1 342.644	719			
2	回归	37.977	2	18.989	10.435	0.000[b]
	残差	1 304.667	717	1.820		
	总计	1 342.644	719			
3	回归	50.929	3	16.976	9.410	0.000[c]
	残差	1 291.716	716	1.804		
	总计	1 342.644	719			

注:a. 预测变量:(常数)、居住地规模(以千人为单位)。

b. 预测变量:(常数)、居住地规模(以千人为单位)、调查对象的年龄。

c. 预测变量:(常数)、居住地规模(以千人为单位)、调查对象的年龄、婚否。

d. 因变量:党派(认为自己是自由党人还是保守党人)。

回归系数[a]

模型		非标准系数		标准系数		
		B	标准误	β	t 值	显著性
1	(常数)	4.280	0.053		81.101	0.000
	居住地规模	0.000	0.000	-0.127	-3.444	0.001
2	(常数)	3.873	0.146		26.463	0.000
	居住地规模	0.000	0.000	-0.124	-3.376	0.001
	调查对象的年龄	0.009	0.003	0.110	2.980	0.003
3	(常数)	4.254	0.204		20.891	0.000
	居住地规模	0.000	0.000	-0.117	-3.170	0.002
	调查对象的年龄	0.009	0.003	0.112	3.056	0.002
	婚否	-0.270	0.101	-0.099	-2.679	0.008

注:a. 因变量:党派(认为自己是自由党人还是保守党人)。

删除变量[d]

模型		β	t	显著性	偏相关	共线性 容忍度
1	受教育年限	−0.082[a]	−2.205	0.028	−0.082	0.996
	年龄	0.110[a]	2.980	0.003	0.111	0.999
	婚否	−0.096[a]	−2.592	0.010	−0.096	0.994
	性别	−0.030[a]	−0.799	0.424	−0.030	0.998
2	受教育年限	−0.057[b]	−1.508	0.132	−0.056	0.932
	婚否	−0.099[b]	−2.679	0.008	−0.100	0.993
	性别	−0.034[b]	−0.928	0.354	−0.035	0.996
3	受教育年限	−0.069[c]	−1.815	0.070	−0.068	0.921
	性别	−0.023[c]	−0.622	0.534	−0.023	0.982

注:a. 保留在模型中的预测变量:(常数)、居住地规模(以千人为单位)。

b. 保留在模型中的预测变量:(常数)、居住地规模(以千人为单位)、调查对象的年龄。

c. 保留在模型中的预测变量:(常数)、居住地规模(以千人为单位)、调查对象的年龄、婚否。

d. 因变量:党派(认为自己是自由党人还是保守党人)。

问　题

1. 哪个(些)是自变量? 哪个(些)是因变量? 它们的测量水平是什么?
2. 所检验的假设是什么?
3. 为什么这里用多元回归统计技术是合适的?
4. 用文字描述这个统计表格所展示的内容。R 和 R^2 分别是什么意思? 模型 1、2、3 分别是什么意思? 你看的是哪一个模型? 哪些自变量是因变量的良好预测变量? B 和 β 有什么不同? 为什么它们的值有些是负的?

5.“删除变量”分表中的数据告诉我们些什么?
6.用日常生活语言表述你所得出的结论。
7.你能否得出结论说,自变量和因变量之间有或者没有关系?还有些什么别的信息可以帮助你作出结论?
8.这些结果的可能解释是什么?有没有理论或者以前的研究可以引导你解读这项研究中的发现?

回归分析小结

在社会科学研究中,很常用的一种数据分析方法就是常最小二乘回归。这种方法通常也称作多元回归,其目的是用两个或者更多自变量来预测因变量的变化情况。该方法必须假定,各自变量和因变量之间的关系是线性的,且自变量之间呈低度相关。

解释多元回归结果的关键是 R(复相关系数)和 R^2(确定系数)的值,以及各个回归系数的大小和方向,包括非标准回归系数 B 和标准回归系数 β。如果用的是逐步回归,所需要的信息通常在每个表格的最后一部分中给出;删除变量部分所列出的变量对预测或者说明因变量都没有统计显著性。

不过,如果整体回归模型没有达到预先设定的统计显著性水平(方差分析结果中的 F 值告诉我们这方面的信息),那么,很可能几个自变量中没有一个能够显著地预测因变量。R^2 值越大,因变量中被自变量解释或预测的变异方差就越多。回归系数值刻画了各个自变量对预测结果贡献的大小。通常,贡献主要是由几个回归系数较大的变量做出的。

参考文献

Auton, Heather R., Jacqueline Pope, and Gus Seeger. 2003. "It Isn't That Strange: Paranormal Belief and Personality Traits." *Social Behavior and Personality* 31:7, 711-720. Table used by permission of *Social Behavior and Personality.*

Baird, Carol L., Donna Schmeiser, and Karen T. Yehle. 2003. "Self-Caring of Women with Osteoarthritis Living at Different Levels of Independence." *Health Care for Women International* 24, 617-634. Table Copyright © 2003 by Baird, et al. and reproduced by permission of Taylor & Francis, Inc. http://www.taylorandfrancis.com.

Bisset, Janet D., Marianne E. Borja, Deborah E. Brassard, Janet R. Reohr, Kathleen O'Neil, and Kathleen O'Neil Ruthkosky. 1999. "Assessing the Importance of Educational Goals: A Comparison of Students, Faculty, and Parents." *Assessment & Evaluation in Higher Education* 24:4, 391-398. Table reproduced by permission of Taylor & Francis, Inc. http://www.tandf.co.uk/journals.

Bornman, Elirea. 1999. "Self-Image and Ethnic Identification in South Africa." *Journal of Social Psychology* 139:4, 411-425. Table reprinted with permission of the Helen Dwight Reid Edu-

cational Foundation. Published by Heldref Publications, 1319 Eighteenth St., NW, Washington, DC 20036-1802. Copyright © 1999.

Brock, Thomas, Isaac Kwakye, Judy C. Polyné, Lashawn Richburg-Hayes, David Seith, Alex Stepick, Carol Dutton Stepick, with Tara Cullen and Sarah Rich. 2004. "Welfare Reform in Miami: Implementation, Effects, and Experiences of Poor Families and Neighborhoods." New York: MDRC. Table used by permission of MDRC (www.mdrc.org).

Burnett, Derek M., Stephanie A. Kolakowsky-Hayner, Joy M. White, and David X. Cifu. 2002. "Impact of Minority Status Following Traumatic Spinal Cord Injury." *NeuroRehabilitation* 17, 187-194. Table reprinted with permission from IOS Press.

Chambers, Sharon M., James C. Hardy, Brenda J. Smith, and Sarah F. Sienty. 2003. "Personality Indicators and Emergency Permit Teachers' Willingness to Embrace Technology." *Journal of Instructional Psychology* 30:3, 185-188. Table published by permission of the *Journal of Instructional Psychology*.

Gfroerer, Kelly P., Coleman A. Gfroerer, William L. Curlette, JoAnna White, and Roy M. Kern. 2003. "Psychological Birth Order and the *Basis-A Inventory*." *Journal of Individual Psychology* 59:1, 30-41. Table used by permission of and copyright © 2003 by the University of Texas Press. All rights reserved.

Grosswald, Blanche. 2003. "Shift Work and Negative Work-to-Family Spillover." *Journal of Sociology and Social Welfare* 30:4, 31-56. Table used by permission of the *Journal of Sociology and Social Welfare*.

Hall-Elston, Claudia, and Larry C. Mullins. 1999. "Social Relationships, Emotional Closeness, and Loneliness Among Older Meal Program Participants." *Social Behavior and Personality*

27:5, 503-518. Table used by permission of *Social Behavior and Personality*.

Huff, Darrell. 1954. *How To Lie With Statistics*. New York: Norton.

Joshi, Anupama, and Jennifer C. Ferris. 2002. "Causal Attributions Regarding Conflicts Between Friends in Middle Childhood." *Social Behavior and Personality* 30:1, 65-74. Table used by permission of *Social Behavior and Personality*.

Kakavoulis, Alexandros. 2001. "Family and Sex Education: A Survey of Parental Attitudes." *Sex Education* 1:2, 163-174. Table reproduced by permission of Taylor & Francis, Inc. http://www.tandf.co.uk/journals.

Kemple, James J. with Judith Scott-Clayton. 2004. "Career Academies: Impacts on Labor Market Outcomes and Educational Attainment." New York: MDRC. Table used by permission of MDRC (www.mdrc.org).

Looker, E. D. and Victor Thiessen. 1999. "Images of Work: Women's Work, Men's Work, Housework." *Canadian Journal of Sociology* 24:2, 225-254. Table used by permission of the *Canadian Journal of Sociology*.

Mayo, Joseph A. 2002. "Case-Based Instruction: A Technique for Increasing Conceptual Application in Introductory Psychology." *Journal of Constructivist Psychology* 15:65-74.

Menet, Fiona, John Eakin, Michael Stuart, and Harry Rafferty. 2000. "Month of Birth and Effect on Literacy, Behaviour and Referral to Psychological Service." *Educational Psychology in Practice* 16:2, 225-234. Table reproduced by permission of Taylor & Francis, Inc. http://www.tandf.co.uk/journals.

Rovai, Alfred P. and Marcus D. Childress. 2002/2003. "Explaining & Predicting Resistance to Computer Anxiety Reduction among Teacher Education Students." *Journal of Research on*

Technology in Education 35:2, 226-235. Table reprinted with permission from the *Journal of Research on Technology in Education*, copyright © 2002/2003, ISTE (International Society for Technology in Education), 1.800.336.5191 (US & Canada) or 1.541.302.3777 (International), iste@iste.org, www.iste.org. All rights reserved.

Salka, William M. 2003. "Determinants of Countywide Voting Behavior on Environmental Ballot Measures: 1990-2000." *Rural Sociology* 68:2, 253-277. Table used with permission of the Rural Sociology Society.

Stein, Robert M. 1998. "Introduction: Early Voting." *Public Opinion Quarterly* 62:1, 57-69. Table used by permission of Oxford University Press.

Stevens, Daphne Pedersen, Gary Kiger, and Pamela J. Riley. "Coming Unglued? Workplace Characteristics, Work Satisfaction, and Family Cohesion." *Social Behavior and Personality* 30:3, 289-302. Table used by permission of *Social Behavior and Personality*.

Trudel, Gilles. 2002. "Sexuality and Marital Life: Results of a Survey." *Journal of Sex & Marital Therapy* 28:229-249. Table Copyright © 2002 by Trudel and reproduced by permission of Taylor & Francis, Inc. http://www.taylorandfrancis.com.

Yoon, Jina S. 2002. "Teacher Characteristics as Predictors of Teacher-Student Relationships: Stress, Negative Affect, and Self-Efficacy." *Social Behavior and Personality* 30:5, 485-494. Table used by permission of *Social Behavior and Personality*.

推荐阅读

本书作者的网页 http://pzacad. pitzer. edu/~pnardi/soc101links. html 上提供了大量的网络资源信息,欢迎大家访问。

Babbie, Earl. 2003. *The Practice of Social Research.* Tenth Edition. Belmont, CA: Wadsworth.

Campbell, Donald and Julian Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research.* Boston: Hoghton Mifflin.

Miller, Delbert C. 1991. *Handbook of Research Design and Social Measurement.* 5th ed. Newbury Park, CA: Sage.

Nardi, Peter M. 2003. *Doing Survey Research: A Guide to Quantitative Methods.* Boston: Allyn & Bacon.

Neuman, W. Lawrence. 2003. *Social Research Methods: Qualitative and Quantitative Approaches.* Fifth Edition. Needham Heights, MA: Allyn & Bacon.

Norusis, Marija A. 2004. *SPSS* 12.0 *Guide to Data Analysis.* New York: Prentice Hall.

Phillips, John L. Jr. 2000. *How To Think About Statistics.* Sixth Edition. New York: Freeman.

附　录

Scale 及其相关术语的翻译说明

在英文社会统计学和社会测量学文献中，scale 这个词不仅自身的使用频率很高，而且还派生了多个计量学词汇，例如 a scale，the scale，scales，scaling 等。由于英汉两种语言在词汇层面的对应性很低，所以这给翻译带来很大的困难。具体而言，scale 一词在英语中所覆盖的内涵非常丰富，目前汉语中还没有一个词汇或术语能够与它对应。对于这种情况，我们认为翻译时，最好根据具体的语境，选择不同的汉语译名，以保障汉语译文的可读性。

英文 scale 一词的意义很广。根据现行美国《教育与心理测试标准》词汇表中的定义，scale 有两个社会计量学专业意思。第一个意思指一个标准化的度量制度，其中包括单位和单位制。“温标”（包括摄氏、华氏、绝对等）就是这个意思在温度测量中的具体化。英文 scaling 一词就是 scale 一词在这个意义上的派生词。在实践中，这一个意思还有一般（即抽象）和具体的区分。一般性意义上的 scale 泛指一切标准化了的量化制度或体系；具体性意义上的 scale 特指某一类标准化了的量化制

度或体系。第二个意思是指一组项目、一个项目集或子集。在这个意义上,scale 与 test 同义,既有点儿"测量工具"或"测量仪器"的意思,也有点"代表性样本"的意思。

在有关汉语文献中,通常用"量表"与英语的 scale 对应。这个广为接受的解决方案是一个过分简单化了的方案,也是一个包含歧义的方案,因此,很容易引起读者的误解。"量表"中的"量"是"测量"意义上的"量"? 还是"量值"、"数量"意义上的"量"? "量表"中的"表"是"表格"、"项目单"意义上的"表"? 还是"钟表"或"温度表"意义上的"表"? 关于这些,文献中都没有具体的定义或说明。即使我们可以假定"量表"中的"量"指"量值"或"数量","量表"中的"表"指"钟表"或"温度表"中的"表","量表"一词所包含的内容的确很多,也可能太多。试想,电表、米尺、水表、里程表、日历、二十四节气等哪个测量工具不是"量表"啊?

考虑到英语 scale 一词意义的广泛性和复杂性以及汉语"量表"一词的歧义性,我们在翻译这本时只好具体问题具体解决。当 scale 一词在一般意义上使用时,我们译作"量标",以揭示意义的一般性。当它指某类量的量化制度时,我们则用"××标"(如,利克特氏态标、级标),以尽量表明所测定的量的性质。当它指某类量的量化制度的物化代表时,我们用"××计"或"××量计"(如,利克特氏态度计),以避开由于"量表"一词引起的含糊和混乱。我们翻译时所遵循的原则是普通计量学对于量的表示的基本要求,即量 = 数值 + 单位。在社会测量中,无论是量的数值范围,还是量的单位,都包含在所用的测量工具之中。因此,表述测量结果时,如果不表明具体的量化制度,不仅所表述的量的数值没有意义,它的单位也没有着落。

译　者

2009 年 5 月

汉英术语对照表

(单向)方差分析	ANOVA
F 检验	F-test
F 值	F value
Levene 氏方差齐性检验	Levene F-test for equality of variance
p 值	p value
SPSS 结果输出	SPSS output
t 检验	t-tests
百分位数	Percentile
边际	Marginals
变量	Variables
变异性	Variability
标准差	Standard deviation
标准化回归系数	Standardized regression coefficient
标准化回归系数	Beta [standardized regression coefficients]
标准误差	Standard error
表格	Tables
饼图	Pie charts
参数	Parameters
常最小二乘线性回归	OLS linear regression/Ordinary Least Squares linear regression
抽样误差	Sampling error
大众出版物	Popular publication
单元格	Cells

第一类错误	Type I error
定比变量	Ratio measured variables
定距/定比变量	Interval/ratio measured variables
定类变量	Nominal measured variables
定序变量	Ordinal measured variables
独立样本 t 检验	Independent-sample t-tests
对第三个变量的控制	Controlling for third variable
多变量交叉表	Multiple crosstab tables
多数人所选中的项	Majority response
多元回归分析	Multiple regression analysis
多元数据	Multiple data
二元表格	Bivariate tables
二元数据	Bivariate data
非标准化回归系数	B [unstandardized regression coefficients]
非标准化回归系数	Unstandardized regression coefficients
复相关	Multiple correlation
复相关系数	R
概括性	Generalizability
回归分析	Regression analysis
假设	Hypothesis
交叉表	Cross-tabulations
均值	Mean
卡方检验(χ^2)	Chi-square (χ^2) test
累积百分数	Cumulative percent
利克特态度级标	Likert scales
两分判断	Dichotomies
量的水平	Level of measurement
量度/测量	Measurement

列	Columns
列联表	Contingency tables
零假设	Null hypothesis
描述数据	Describing data
描述统计	Descriptive statistics
配对样本 t 检验	Paired-sample t-tests
皮尔逊相关系数 r	Pearson r correlation coefficient
频数表	Frequency tables
频数多边图	Frequency polygons
趋中量	Measures of central tendency
缺失数据以应答者为单位删除	Listwise deletion
决定系数	R^2 (coefficient of determination)
人口统计学属性	Demographics
深入分析	Elaboration
数据	Data
斯皮尔曼相关	Spearman correlation
四分位数	Quartiles
随机抽样	Random selection
条形图	Bar charts/graphs
统计	Statistics
统计分析	Statistical analysis
统计显著性	Statistical significance
推断统计	Inferential statistics
显著性	Significance
显著性水平	Significance level
线性回归	Linear regression
相关	Correlation
协方差	Covariance
行	Rows
学术论文	Scholarly articles

样本	Sample
一次性进入式回归分析	Enter method regression analysis
一元数据	Univariate data
因变量	Dependent variables
有效百分数	Valid percent
预测变量	Predictor variables
折线图	Line graphs
直方图	Histograms
值	Value
中位数	Median
众数	Mode
逐步回归法	Stepwise method
逐步回归分析	Stepwise method regression analysis
准则变量	Criterion variables
自变量	Independent variables
总体	Population
组	Groups

《怎样用数据说谎》 *How to Lie with Statistics*